AD A119218

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 82-0714	2. GOVT ACCESSION NO.: A119218
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): PHYSIOLOGICAL AND DUAL TASK ASSESSMENT OF WORKLOAD DURING TRACKING AND SIMULATED FLIGHT	5. TYPE OF REPORT & PERIOD COVERED: Final Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Dr. Ernest Lindholm	8. CONTRACT OR GRANT NUMBER(s): F49620-79-C-0197
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Psychology, Arizona State University, Tempe, Arizona 85287	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2313/A2
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NL, Bolling AFB, DC 20332	12. REPORT DATE: January 1981
	13. NUMBER OF PAGES: 77
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE SEP 14 1982 D H

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Heart Rate; Event Related Potentials; Evoked Potentials, Skin Conductance; Dual Task; Reaction Time; Information Processing; Workload; Simulated Flight; Tracking

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A visuomotor task of moderate complexity (tracking) and one of high complexity (simulated aircraft carrier landing) were performed alone, then in combination with a tone discrimination task at two levels of difficulty in usual dual task fashion. Measures of autonomic nervous system activation (heart rate, skin conductance) and central nervous system information processing (event related potentials) were quantified continuously during performance of all tasks. The dual task results were typical, given that most subjects treated the tone discrimination task as "secondary" (low priority):

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

tone discrimination performance degraded when the tone task was combined with the tracking task and degraded even more when the tone task was combined with the carrier landing task. While dual task methodology adequately described gross changes in workload, the physiological data permitted much more detailed interpretations and descriptions of training effects (practice), tone task information processing, individual differences, and visuomotor task control parameters than was possible by analysis of secondary task performance. It is concluded that the physiological method has distinct advantages over the dual task method, due mostly to the nonintrusive nature and the greater detail of results afforded by the former method.

AFOSR-TR- 82-0714

Title: Physiological and Dual Task Assessment of Workload During Tracking and Simulated Flight.

Authors: Ernest Lindholm, Cary Cheatham, and George Buckland

Performing Organization: Arizona State University,
Department of Psychology,
Tempe, Arizona 85287

Controlling Office: AFOSR
Bolling AFB
D.C. 20332

Contract Number: F49620-79-C-0197

Type of Report and Period Covered: FINAL

Key Words:

heart rate	reaction time
event related potentials	information processing
evoked potentials	workload
skin conductance	simulated flight
dual task	tracking

ABSTRACT

A visuomotor task of moderate complexity (tracking) and one of high complexity (simulated aircraft carrier landing) were performed alone, then in combination with a tone discrimination task at two levels of difficulty in usual dual task fashion. Measures of autonomic nervous system activation (heart rate, skin conductance) and central nervous system information processing (event related potentials) were quantified continuously during performance of all tasks. The dual task results were typical, given that most subjects treated the tone discrimination task as "secondary" (low priority): tone discrimination perfor-

mance degraded when the tone task was combined with the tracking task and degraded even more when the tone task was combined with the carrier landing task. While dual task methodology adequately described gross changes in workload, the physiological data permitted much more detailed interpretations and descriptions of training effects (practice), tone task information processing, individual differences, and visuomotor task control parameters than was possible by analysis of secondary task performance. It is concluded that the physiological method has distinct advantages over the dual task method, due mostly to the nonintrusive nature and the greater detail of results afforded by the former method.

I. INTRODUCTION

In complex human-machine systems, one important determinant of performance is the ability of the operator to perform multiple tasks. An aircraft pilot, for example, must visually process information from the instrument panel and surrounding environment, respond to some auditory messages received over the headset and not respond to others, and manipulate a number of control devices all more or less simultaneously. As aircraft have become faster and more responsive over the years, demands on the pilot to respond rapidly and accurately have obviously increased; also, in combat or other unusually dangerous situations, emotional stress can seriously interfere with the operator's ability to make optimal decisions, thus complicating the situation even more.

For these reasons, it is convenient to view the pilot as a highly trained and specialized biological system whose task is to receive relevant information through sensory channels, make correct decisions very rapidly, and translate these decisions into patterns of motor activity which result in optimal control of the aircraft. Obviously, there can be situations in which the cognitive demands (mental workload) on the pilot exceed his or her ability to cope, and gross performance errors will result. One role of the psychologist is to analyze the effects of workload on performance with the aim of improving performance and reducing the probability of gross errors. This requires ways to study workload-performance relationships in situations which pose no actual threat to life and property while, at the same time, approximating real-life situations as closely as possible.

Several methods for assessing workload have been developed over the years, but all have limitations and drawbacks (Wierwille and Williges, 1978). This is so

for two primary reasons, a) the phenomena under study are exceedingly complex, and b) in order to determine relationships among workload and performance, the experimenter must create some conditions in which workload is excessive and performance decrements are clear, yet this must be done with the restriction that life and property are not threatened. This leads to laboratory analogs of actual flight which lack realism to varying degrees.

What is needed are methods of workload assessment which deal with a) as well as possible but eliminate the requirement of b), that workload be made excessive. Of the methods currently available, physiological assessment appears the most adaptable to meeting these needs. While other methods vary workload by increasing the number and/or difficulty of multiple tasks until performance decrements are observed, the physiological method can be used to monitor the internal state of the operator under normal task demands. It is convenient to refer to the former methods as "intrusive" since task load is purposely increased until it intrudes upon the operator's ability to perform well. A related problem is that multiple tasks require multiple motor responses, so it is frequently not clear whether a performance decrement is due to high mental workload or response interference (McLeod, 1978).

In this context, physiological assessment can be viewed as nonintrusive since no responses other than those normally emitted by the operator are required. Workload is quantified, not in terms of secondary task reaction times or error scores dependent on skeletal motor responses, but in terms of autonomic and central nervous system responses which reflect variations in physiological function introduced by variations in workload. The sensitivity of physiological responses to variations in mental workload has been demonstrated in a

variety of tasks including mental arithmetic (Ahern and Beatty, 1979; Kahneman, Tursky, Shapiro, and Crider, 1969), psychophysical judgements (Lang, Gatchel, and Simons, 1975), and common laboratory information processing tasks such as choice reaction time and letter matching (Lindholm, Ruppel, and Buckland, 1979). Heart rate, skin conductance, and pupil size are reported to be consistently related to workload, but pupil size has the disadvantage of being very difficult to quantify in environments where the head is free to move. A central nervous system response of great promise in workload studies is the event related potential recorded from the surface of the scalp. Lindholm et al., (1979) reported that components of the event related potential occurring as early as 200 msec following stimulus onset discriminated task difficulty and performance in choice reaction time and letter matching tasks. Isreal, Wickens, Chesney, Donchin (1981) found that the amplitude of a particular component, the P300, changed with workload in a task involving the monitoring of a simulated air traffic control display.

One limitation of the above work is that the tasks used are simple and tend to be poor laboratory analogs of real situations (the Isreal et al., report might be considered an exception), thus it remains to be demonstrated that physiological assessment has utility in more complex situations. Another limitation is that most experiments have quantified only one physiological variable at a time, and it is unlikely that phenomena as complex as mental workload can ever be satisfactorily measured or estimated by a single variable. The Lindholm et al., study did use multiple physiological measures, but the tasks were not complex.

This report describes the results of a 30 month effort which is a logical con-

tinuation of earlier work (Lindholm et al., 1979). As before, multiple physiological variables are quantified, but the tasks were carefully chosen to form a close analog to real world situations. The major task is a computer simulation of landing a Navy A7 aircraft on an aircraft carrier. To simulate the pilot task of processing auditory information received over the headset, subjects were required to perform a tone discrimination task either alone or in combination with the aircraft landing task. Also, another visuomotor control task (continuous tracking) much simpler than the aircraft landing task was employed, and subjects performed this task alone or in combination with the tone discrimination task. In this manner, both dual task and single task paradigms are represented so that the ability of physiological measures to describe workload can be analyzed in both paradigms.

II. METHOD

Subjects

Eight males (ages 20-22 years) were recruited from the on-campus Air Force ROTC program. None had ever piloted a jet aircraft. Two subjects did not complete the experimental procedure and their data were not analyzed. The remaining 6 subjects are designated by the letter codes B,C,D,E,F, and H.

Apparatus

Computer and peripherals- A digital Equipment Corporation PDP 11/34a computer with 112,000 words of MOS memory was programmed to control all phases of the experiment. Important peripherals included a Digital Equipment Corporation VT-11 video display, an ADAC 12 bit 16 channel analog to digital (A/D) converter and a 2 channel digital to analog (D/A) converter. Pertec and Control Data Corporation magnetic disk systems were used for on-line storage, and Pertec magnetic tape systems for off-line data storage. Tone stimuli were

generated by a Wavetek voltage controlled oscillator which was driven by one of the computer D/A outputs. A Beckman Type 411 6-Channel Dynograph was used to amplify and condition all physiological signals; the high level outputs of the Dynograph served as inputs to the A/D converters.

Subject booth- The VT-11 display was placed on a shelf 68 cm above the floor of an electrically shielded booth the inside dimensions of which were 1.2 meters long by 0.8 meter wide by 1.7 meters high. The booth also contained a chair in which the subject sat, and affixed to the chair was a full-sized gimbol joystick of the type found in older multi-engine aircraft. The joystick was modified with small gears and shafts so that movements of the joystick rotated the shafts of two potentiometers, one for left-right movements and one for back-forth movements. Lantern batteries were connected across the potentiometers and the voltage outputs from the wipers were fed to two channels of the A/D converter. In this manner, the full range of possible stick movements was translated into digital information and made available to the software control programs. The joystick was mounted in the center and just in front of the chair; the subject sat with one leg on either side of the stick and grasped the stick with his right hand. A throttle was mounted on the left side of the chair but this was inoperative in the present experiments and subjects were so informed.

Description of Tasks and Scoring Schemes

Tone Discrimination Task- a standard tone of 1500 Hz was presented 10 times at a repetition rate of once per 3 seconds. Thirty to 45 sec later, a series of 24 comparison tones was presented at a repetition rate of once per 5 seconds. The 24 comparison tones consisted of 6 repetitions of 4 tones (1000, 1250,

1750, and 2000 Hz) presented in random order. In all cases, tone duration was 200 msecs and loudness was 65 dB. Subjects were instructed to listen to the reference tone series, then to respond as rapidly as possible to the comparison tones that were either higher than the reference tone (respond high condition) or lower than the reference tone (respond low condition). The response consisted of saying the word "tone" into a lapel microphone which was fixed in a harmonica brace worn about the subject's neck. The microphone was adjusted to within 2 inches of the subject's lips. Microphone output was amplified by one channel of the Dynograph and digitized by one channel of the A/D converter. A software routine monitored this channel and measured reaction times to the nearest 4 msecs. Failure to respond within 1.5 sec of tone onset was scored as an error of omission.

Tracking Task- The PDP 11/34 was programmed to present, on the VT-11 display, an octagon with vertical and horizontal sides of 6.9 cm and angular sides of 8.4 cm as the path to be tracked. Also programmed was a diamond-shaped "bug" with sides of 0.8 cm which could be "flown" anywhere on the VT-11 screen by appropriate joystick movements. Pulling back on the stick propelled the bug toward the top of the screen, and forward movements of the stick produced the opposite result. Left and right stick movements produced compatible bug movements, and angular stick movements produced veridical bug movements (e.g., pushing the stick forward and to the right would cause the bug to move simultaneously to the right and toward the bottom of the screen). Rate of bug movement was a monotonic function of stick displacement. In this manner, the subject could propel the bug around the octagon path with as much speed and accuracy as his individual talents permitted.

The VT-11 also displayed, in the center of the octagon path, two feedback variables: 1) Time Remaining, counted down from 120 sec in one sec decrements, and 2) Laps Completed, which incremented from zero in 1/4 lap increments (a lap was defined as one complete trip around the octagon). Subjects were instructed to fly the bug around the octagon in a clockwise direction as rapidly as possible while, at the same time, staying as close to the path as possible. They were told that their score depended on both accuracy and speed. The score was computed, on line, according to the following formula: Score= 100 - (100(z)/(z(M)), where z is the total root mean square deviation of the bug from the octagon path and M is the number of x-y measurements per 2 min run. Since a measurement was taken whenever the bug was displaced a fixed distance (1 cm) in either the x or y plane, M becomes a measure of speed of bug movement.

Carrier Landing Task- This is a complex software package developed for the PDP 11/34 and VT-11 display by personnel at NTEC, Orlando, Florida. Flight dynamics approximate those of the Navy A7 aircraft. Displayed on the upper 2/3 of the VT-11 is a full length horizon and an aircraft carrier with a wake. Carrier simulation includes a rudimentary superstructure, bow, stern, waterline, landing area with centerline, and the Fresnel Optical Landing System (FOLS). The latter consists of two short horizontal lines located just above the carrier deck and to the left of the landing area. A small ball located between the lines moves above the lines if the pilot is too high on approach, and below the lines if the pilot is too low on approach. The FOLS (also called the "meatball") is clearly visible from a simulated 4 miles aircraft-to-carrier distance and becomes more clear as distance decreases.

Displayed on the lower portion of the VT-11 are several cockpit instruments,

specifically in these experiments, altimeter, vertical speed indicator, attitude indicator, compass, TACAN (station on carrier), and percent engine power. The throttle was frozen at 87% power in these experiments to simplify the task and free the left hand for skin conductance measurements. All cockpit instruments, as well as the out-the-window display of the horizon, carrier, and FCLS, changed in real time as functions of joystick manipulation. Refresh rate was 30 Hz.

Software routines detected, computed, and reported, on line, the following situations, any one of which froze the display and terminated the flight:

1) *Splash- aircraft impacted with water* (altitude reached zero feet. Since the carrier deck had a fixed altitude of 60 feet, splash was not confused with any of the other situations described below).

2) Crash- Aircraft was over the landing area of the carrier, but either roll was excessive (greater than 10 degrees, indicating that wingtip struck carrier deck) and/or vertical speed was excessive (decent rate of greater than 2000 feet per minute was scored as a nosedive into the carrier deck).

3) Bolter- almost a landing, but aircraft attitude was not within limits to catch one of the four available tail-hook wires. This would occur when the pilot "bounced" on the deck due to excessive vertical speed (greater than 1000 but less than 2000 feet per minute).

4) Ramp Strike- approach was too low and aircraft struck the stern of the carrier just below the landing area.

5) Timeout- Pilot got lost, flew in the wrong direction, or missed the carrier on approach. Timeout occurred after 2-1/2 min of flight; a reasonable approach and landing required 2 min.

6) Landing- This was the goal of this particular task. A landing meant that

the subject performed reasonably well during the approach (thereby avoiding a splash or timeout) and also made contact with the landing area in a smooth fashion (thereby avoiding a ramp strike, bolter or crash).

The subjects were released from freeze under the following conditions: altitude of 1550 feet above sea level, heading of 351°, distance to carrier of 3.8 nautical miles, flaps full and gear down, vertical speed of -200 feet per minute, and a constant 87% power, which resulted in an initial airspeed of 125 knots. With constant percent power, airspeed varied slightly as a function of aircraft attitude; climbing would reduce airspeed and diving would increase airspeed. These fluctuations were not great and are considered unimportant to the correct performance of the task. The initial heading of 351° was "ideal" for a straight-in approach; that is, no turns were necessary to line up on the carrier landing area, and from this it follows that ideal roll should be zero degrees. Given the starting altitude and distance to carrier, the ideal descent rate was calculated to be approximately 750 feet per minute vertical speed, and subjects were so informed.

Scores on this task were divided into two subsets, a total approach score and a landing score. The approach score was calculated from three flight parameters, a) root mean square deviations (RMSE) of actual heading from the ideal heading of aircraft to carrier, b) RMSE of actual vertical speed from ideal vertical speed, and c) RMSE of actual roll from ideal roll. The RMSE's were then subtracted from 100 so that subjects could be told simply that a score of 100 was perfect. Negative scores were, of course, possible in this scheme and were recorded and used in statistical analyses; however, scores reported to subjects as feedback were restricted to the range of zero to 100. The second subset,

total landing score, was derived from two parameters, a) lateral offset, in feet, of the aircraft nose from the landing strip centerline at the time the landing was made; this value was subtracted from 100 so again, a score of 100 was perfect, and b) which of the 4 tail hook wires was caught. Wires one and two are nearest the stern and catching these indicated that the approach was lower than optimal, just above a ramp strike; either of these scored 67 on the 100 point scale. Wire 3 is optimal and scored 100 while wire 4 indicated a higher than optimal approach (close to a bolter) and scored 67.

The rationale for this scoring scheme was simple: Flying straight toward the carrier with the proper rate of descent indicated that the subject had the aircraft well under control and this earned a high score. Conversely, large deviations on any of these measures indicated poor aircraft control and this earned a low score. If the flight terminated in a splash, ramp strike, time-out, or crash, a 50 point penalty was subtracted from the approach score. A bolter resulted in only a 10 point penalty since the subject did manage to touch the carrier deck without a crash. In this manner, highest scores were obtained by a smooth approach combined with a landing. At the other extreme, a splash soon after release from freeze would earn a score of zero or less. All intermediate forms of performance were reflected by a large range of possible scores which, we believe, represent an interval scale.

Procedure

Each subject served for 10 hrs and was paid $30. The 10 hrs were distributed approximately equally over three days, each day separated by 4-8 days.

Day 1- the subject was greeted, shown the laboratory, and the three tasks were explained. Electrodes were placed at the following locations while the experi-

menters explained what the electrodes were for and encouraged questions from the subject: a) vertex (Cz) referenced to right mastoid, left mastoid ground, b) lateral canthus and superior ridge of left eye, for eye movement and blinks, c) middle finger and back of left hand for skin conductance, d) sternum and left lateral rib for heart rate. Beckman silver-silver chloride electrodes were used for all placements; the vertex lead was held with Grass electrode paste and a gauze sponge, while all other leads utilized the Beckman double adhesive collars and Beckman electrode cream. Electrode impedance (measured at 30 Hz) was typically less than 5 K-ohms for the vertex and mastoids and less than 30 K-ohms for the other leads. Dynograph bandpass was set at 5.3 Hz to 30 Hz for the eye and heart channels, 5.3 Hz to maximum for the voice channel, DC to 30 Hz for the skin conductance channel, and .16 to 30 Hz for the vertex channel.

The subject wore a light weight junction box around the neck which served to connect all primary leads to the Dynograph through a connector. Thus, the subject could disconnect during breaks and move about (visit rest room, get a drink, stretch their legs).

For all subjects, the first task performed on day 1 was the tone discrimination task. There were 10 runs of 24 tone trials, 5 respond high and 5 respond low. The order of respond high and respond low was the same for all subjects (H,L,L,H,H,L,H,L,L,H) and each run was preceded by the 10 presentations of the reference tone. Following a 10 min break, the second task (tracking task) was demonstrated by one of the experimenters. Subjects were told that they would still hear the tones while they were performing the tracking task, but they did not have to respond to the tones, indeed, they could ignore the tones com-

pletely. Their immediate goal was to earn the highest scores possible on the tracking task. Two practice runs were administered during which no data were collected, then 10-2 min runs were administered with an inter-run interval of 30-45 sec.

The final session on day 1, which began after another 10 min break, was a combination of tasks 1 and 2; that is, subjects were instructed to perform the tracking task while, simultaneously, responding to tones either higher or lower than the reference tone. The sequence of events in this final session of day 1 was as follows: a) Presentation of reference tones in usual manner, b) the subject was told to respond to tones either higher or lower than the reference tone and perform the tracking task simultaneously, c) 2 min of combined task, d) 45 sec inter-run interval. With this sequence, 5 respond high and 5 respond low tone discrimination runs were administered simultaneously with 10 runs of the tracking task. At the conclusion of each run, the subject was told his tracking score and his tone error score. They were consistently encouraged to strive for zero errors on the tone task and simultaneously, highest possible scores on the tracking task. If a subject appeared to be responding slowly,he was reminded that he must respond to the tone within 1.5 sec of tone onset.

Day 1 was concluded by removing electrodes (after recording terminal impedence values), answering any questions the subject might have, and reminding the subject of his day 2 appointment.

Day 2- Electrodes were attached as for day 1 and the carrier landing task was demonstrated with an explanation of the cockpit instruments, the FOLS, and simple strategies for performing the task (e.g., "Make frequent and small stick corrections. Try to keep lined up on the center line of the carrier landing

area.") Subjects were told that, of the cockpit instruments, the altimeter and vertical speed indicator were the most critical. Low altimeter readings warned of impending splash, and the vertical speed indicator should be maintained between -500 and -1000 feet per minute for an ideal approach. They were further told that maintenance of the proper heading was best accomplished by keeping the carrier visually lined up "out the window" in the release-from-freeze relationship, since they were released from freeze with the ideal heading for a straight-in approach. Subjects were reminded that they would hear the tones from the tone discrimination task while they were flying the simulator, but that they were not to respond to the tones on this day. Each subject was then allowed to fly the simulator for 30 runs with a 10 min break after each session of 10 runs. The subject was told his score after each run and the experimenter provided simple advice (e.g., "You came in too high (or too low) that time.", or "Don't forget to watch the meatball carefully as you get close to the carrier.") to encourage better performance.

Day 3- This was identical to day 2 except that subjects were told that they must perform the carrier landing task and the tone discrimination task concurrently. Again, 30 runs were administered with a 10 min break after each session of 10. Within each session of 10 runs, there were 5 respond high and 5 respond low tone discrimination runs. After each run, subjects were told their flight score as well as the number of errors on the tone discrimination task for that run. As in the combined tracking and tone discrimination task, subjects heard the reference tone before each flight.

General- Note that each of the three tasks have run durations of 2 min so when the tone discrimination task is combined with either of the two visuomotor

tasks, the 24 tone trials are equally spaced in time (5 sec inter-tone interval) throughout the 2 min runs of the visuomotor tasks. Also, subjects were given considerable feedback: Error scores on the tone task and performance scores on the tracking and carrier landing tasks were reported to the subject at the conclusion of each 2 min run. Subjects were consistently encouraged to perform as well as possible on all tasks alone and in combination.

III. Results: Among-subjects effects

Tracking and carrier landing task performance

Figures 1a and 1b summarize the performance, averaged over the 6 subjects, on the two visuomotor tasks when they were performed alone and in conjunction with the tone discrimination task. Repeated measures ANOVA's were performed separately for each of the four functions shown. As suggested by inspection of Figure 1a, the runs effect was significant for the tracking task performed alone ($F(9,45) = 8.39$, $p < 0.001$) which simply means that the tracking task was characterized by a learning function of steep slope. When the tracking task was combined with the tone discrimination task, the runs effect was not significant ($F(9,45)=1.41$, $p > 0.20$) indicating that combining the tone task with the tracking task did not produce performance decrements on the tracking task.

Similar results were found for the data presented in Figure 1b. There were 3 sessions of 10 runs each. The sessions effect was significant but did not interact with runs. The sessions effect is redundant with the runs effect since both reflect practice, so to simplify the figure, runs are averaged over

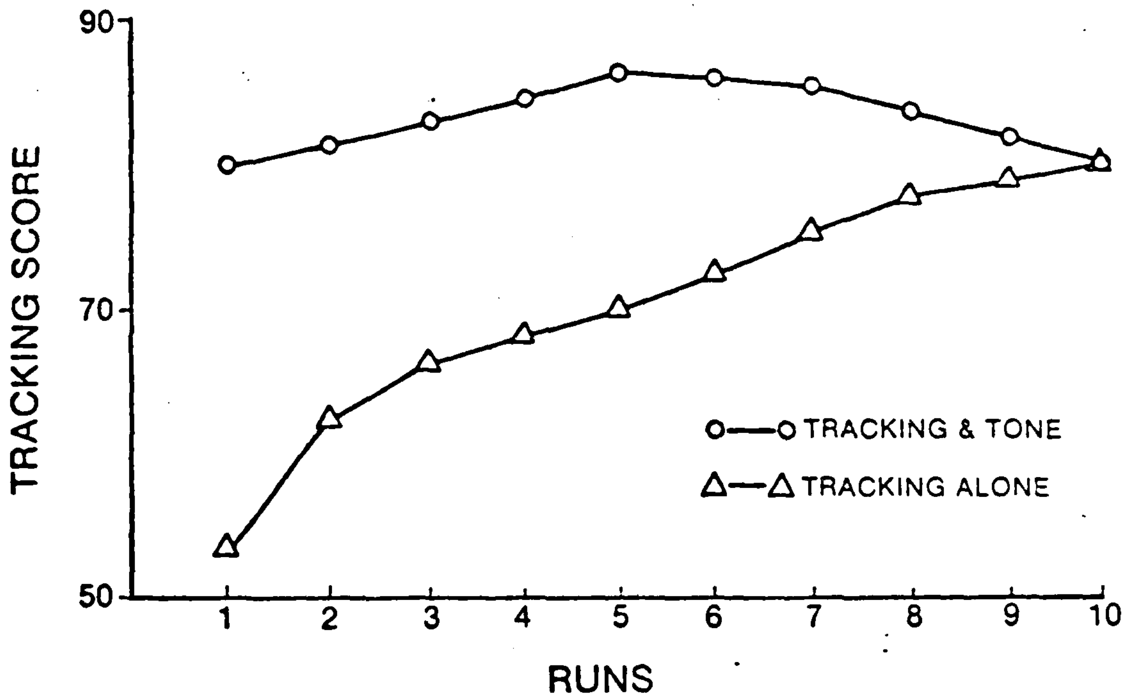

Figure 1a. Performance on the tracking task performed alone and in combination with the tone discrimination task averaged over 6 subjects.

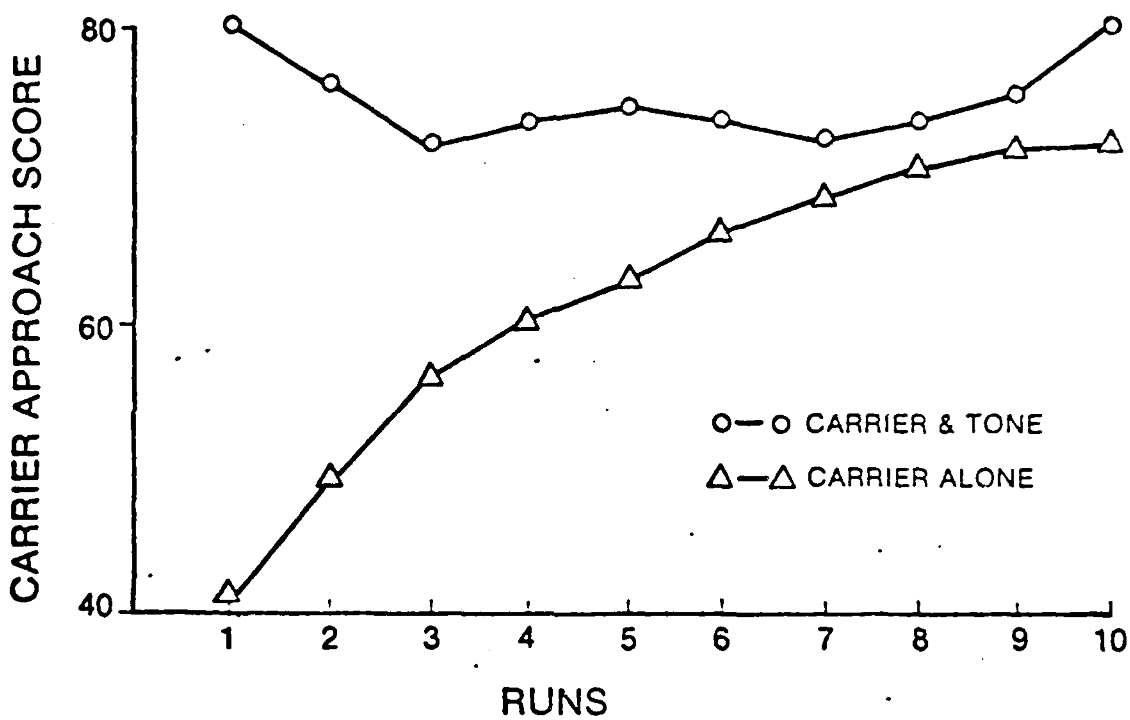

Figure 1b. Performance on the approach to landing segment of the carrier landing task performed alone and in combination with the tone discrimination task averaged over 6 subjects.

sessions. As was the case with the tracking task, the carrier landing task was characterized by a learning function of steep slope (runs for carrier alone ($F(9,45)=7.89$, $P < 0.001$), and combining the tone task with the carrier landing task did not produce performance decrements on the carrier landing task ($F(9,45)=1.15$, $p > 0.3$). Another way to assess performance on the carrier landing is to examine the percentage of successful landings, relative to other ways in which the flight could terminate. These results are shown in Table 1. Splashes and time-outs are combined since both represent the inability of the subject to guide the aircraft to the carrier and thus represent the poorest level of performance. Ramp strikes and crashes indicate that the subject did manage to contact the carrier, but lacked the degree of control necessary to successfully contact the landing area. Bolters occurred when the landing area was contacted, but not within the limits necessary for a successful landing. In this sense, a bolter is an indication of good performance relative to the alternatives of splash, time-out, ramp strike, or crash. Considering first the results for the carrier landing task alone (left half of Table 1), the percentage of splash and time-out decreased over sessions while the percentage of bolters and landings increased. The increase, over sessions, of bolters and landings combined was significant, $F(2,5)=7.61$, $p < 0.05$ as was the decrease in splashes and time-outs ($F(2,5)=6.02$, $P < 0.05$). When the carrier task was combined with the tone task (right half of Table 1), percentage of landings continued to increase so that approximately 50% of all flights during the last two sessions terminated in a successful landing.

To summarize, both visuomotor tasks were characterized by rapid acquisition functions, and performance on these tasks was not degraded by the addition of the tone discrimination task. However, as shown in the following section, performance on the tone discrimination task degraded sharply in the combined task

TABLE 1

Percentage of flights terminated by the methods indicated. Left half of the table shows the results for the carrier landing task performed alone, right half, carrier task combined with tone task. Mean of 6 subjects.

	Carrier Alone			Carrier plus Tone Task		
	Session			Session		
	1	2	3	1	2	3
Splash & Time-out	32	8	2	9	0	2
Ramp Strike & Crash	20	15	15	13	0	8
Bolter	37	58	57	43	50	38
Landing	11	19	26	35	50	52

sessions. Apparently, subjects treated the tone discrimination task as low priority in spite of instructions that both tasks should be performed equally well.

Reaction times and errors for the tone discrimination task performed alone and in combination with the two visuomotor tasks

Preliminary analyses showed that, as expected, neither RT's nor error rates were affected by response set (respond high or respond low), thus the reported results are averaged over this variable.

A repeated measures ANOVA was performed on RT scores with 3 levels of tasks (tone discrimination task performed alone, tracking and tone tasks together, and carrier landing and tone tasks performed together), 2 levels of tone discrimination difficulty (easy versus hard), 2 levels of runs (first 5 runs and last 5 runs), and 4 levels of trial blocks (each block was the average of 6 trials on which the subject was supposed to react with a voice response). Easy tone discriminations involved the two tones furthest from the reference tone frequency while hard discriminations involved the two tones closest to the reference. All 4 main effects and none of the interactions were significant. The mean RT to all tones was 607 msec for the tone discrimination task performed alone, 765 msec for the tracking and tone task combined, and 872 msec for the carrier landing task and the tone task combined ($F(2,10)=39.71$, $p < 0.001$). Mean RT was shorter for the easy discrimination than for the hard (702 msec versus 795 msec, $F(1,5)=50.20$, $p < 0.001$), and mean RT was longer on the second set of runs than on the first set of runs (772 msec versus 725 msec, $F(1,5)=17.24$, $p < 0.01$). Finally, mean RT increased over trial blocks (738

729, 755 and 772 for blocks 1 through 4, respectively, $F(3,15)=4.27$, $p < 0.025$).

Figure 2a summarizes the results of this analysis. To simplify this figure, the runs and easy-hard main effects have been averaged (since there were no interactions this does not misrepresent the functions). It is clear from this figure that the blocks effect, although statistically reliable, is not of impressive magnitude. Recall, however, that subjects must respond within 1.5 sec of tone onset, otherwise the trial would be scored as an error of omission. Thus, the error analysis provides important additional information.

An ANOVA identical to the one described above was performed on the error scores, and the results are displayed in Figure 2b. All main effects excepting runs were significant, and the task by blocks interaction was significant. In agreement with the RT analysis, the main effect of task was significant ($F(2,10)=14.44$, $p < 0.002$), the hard discrimination led to more errors than the easy discrimination ($F(1,5)=78.41$, $p < 0.001$), and errors increased over trial blocks ($F(3,15)=4.28$, $p < 0.025$). The significant task by blocks interaction ($F(6,30)=4.32$, $P < 0.005$) reflects the fact that errors increased more as a function of trial blocks for the carrier plus tone combination than for the tracking plus tone combination or the tone discrimination task performed alone. Subsequent tests (ANOVA's performed on pairs of tasks) confirmed the expectations gained from inspection of Figures 2a and 2b: RT's and errors were greatest when the tone discrimination task was combined with the carrier landing task, intermediate when the tone task was combined with the tracking task, and least when the tone discrimination task was performed alone (all p's < 0.03).

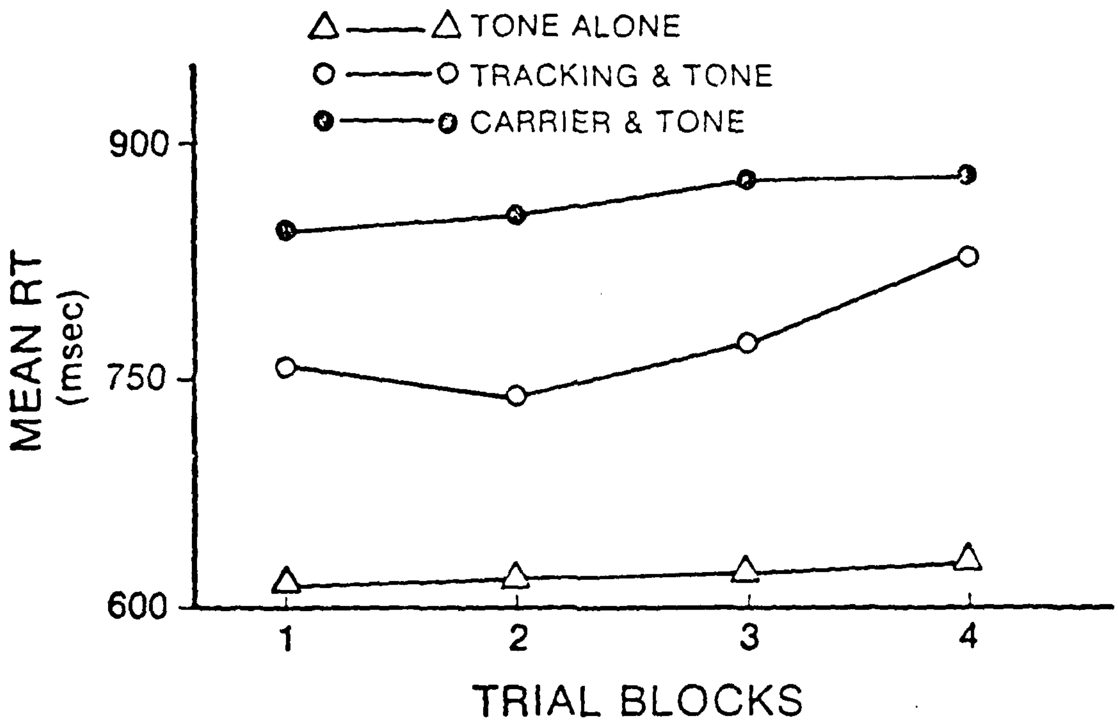

Figure 2a. Mean reaction time as a function of trial blocks for the tone discrimination task performed alone and in combination with each of the visuomotor tasks averaged over 6 subjects.

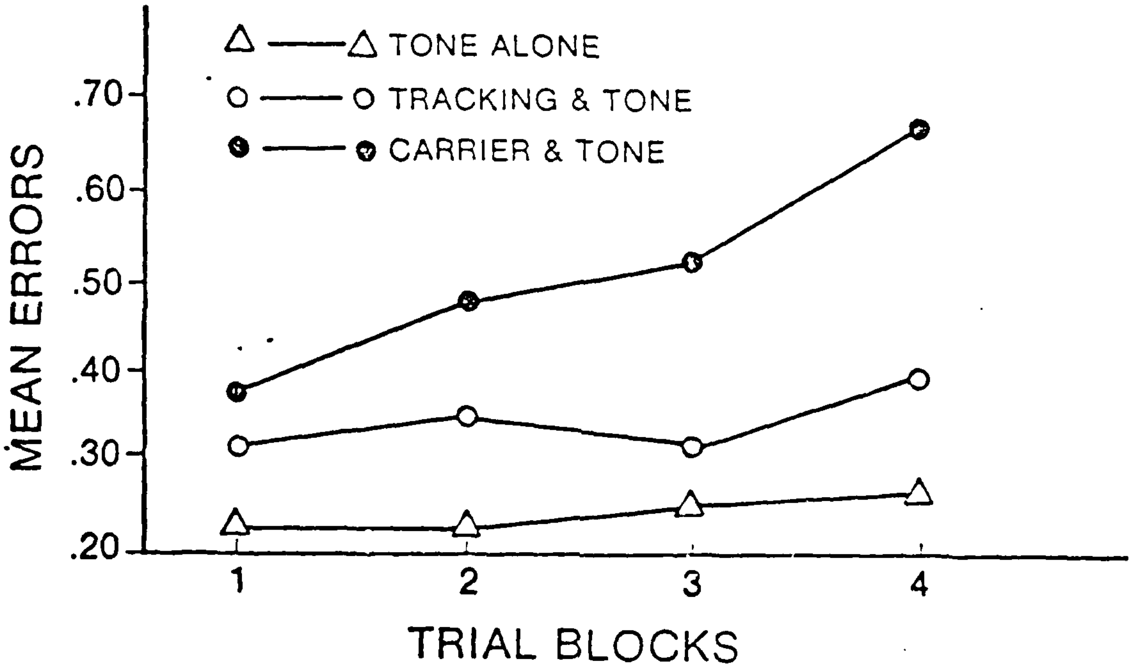

Figure 2b. Mean errors as a function of trial blocks for the tone discrimination task performed alone and in combination with each of the visuomotor tasks averaged over 5 subjects.

In the combined task sessions, RT's were longer on the second set of runs than on the first set of runs, yet primary task performance was stable over runs for the tracking task, and as shown in Table 1, performance continued to increase for the carrier landing task. Thus, the RT results would suggest that workload was increasing with practice, which is unlikely. What seems more probable is that subjects were learning to accurately judge the 1.5 sec time interval during which the RT must be made to avoid an error and also learning that there was no negative reinforcement for long RT's, provided that the RT was less than 1.5 sec. In short, they could treat the secondary task as low priority without reprimand.

The trial blocks effects were particularly interesting since they paralleled the workload gradient over trials which differed for the three tasks. That is, the tone task did not vary in workload as a function of trials since the subject simply heard a tone every 5 sec and had to judge its pitch. However, workload on the tracking task would be expected to increase as a function of trials since the subject could see the number of laps completed and the time remaining, thus subjects would be expected to work harder toward the ends of runs in order to increase their score on the tracking task. Finally, the carrier landing task was very clearly graded in workload as a function of trials; the closer the subject flew to the carrier, the more critical became his processing of the visual display and his stick movements. Indeed, some of our subjects volunteered their opinions that they "simply did not hear the tone" during the last 30 sec of the flight because they were concentrating so much on the final approach to landing. The last 30 sec would correspond to trial block 4 in Figure 2b when errors were highest.

To summarize, when performed alone, the tone discrimination task was characterized by short RT's and low error rates. Errors and RT's increased when the tone task was combined with the tracking task and increased even more when the tone task was combined with the carrier landing task. Within tasks, errors on the tone task increased toward the end of each 2-min run of the visuomotor task, and this effect was pronounced in the carrier landing task as the subject approached the carrier landing area (final approach to landing). Generally, then, RT's and errors on the secondary task increased with increasing primary task demands, as might be expected. One anomalous result was that performance on the tone task became worse as a function of practice on the visuomotor tasks. This is the reverse of what would be expected since practice led to increased mastery and therefore, presumably lower workload. Apparently, with practice, subjects lowered the priority of the tone task in violation of the experimenters' instructions.

Inter-beat interval

Gross body movements would occasionally cause the software trigger to miss an R-wave or mistakenly trigger on the following T-wave, therefore IBI's less than 400 msec or greater than 1500 msec were discarded. Also, preliminary analyses showed that IBI's were not affected by response set (respond high or respond low), thus the data were averaged over this variable.

The IBI's as functions of days and trial blocks are presented in Figures 3a, 3b, and 3c averaged over the 6 subjects. Repeated measures ANOVA's revealed that, on day 1, the only significant effect was the tasks by blocks interaction ($F(10,50) = 3.68$, $p < 0.001$). Inspection of Figure 3a indicates that this effect is due to the fact that IBI's changed very little over blocks during

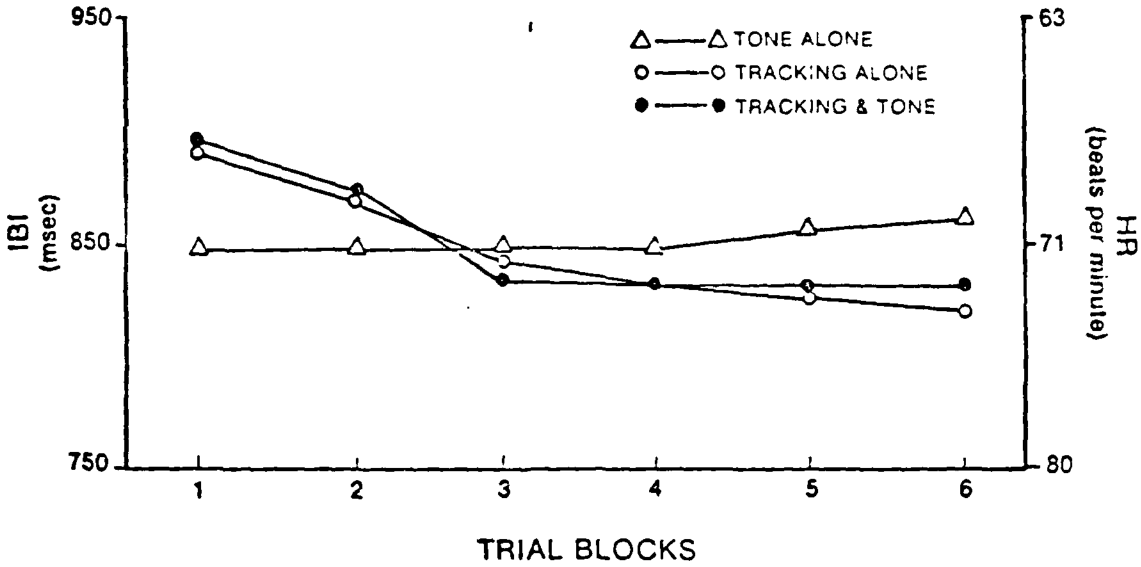

Figure 3a. Mean IBI as a function of trial blocks for the tone discrimination task performed alone, the tracking task performed alone, and the combined task conditions. Mean of 6 subjects. Heart rate plotted on right ordinate.

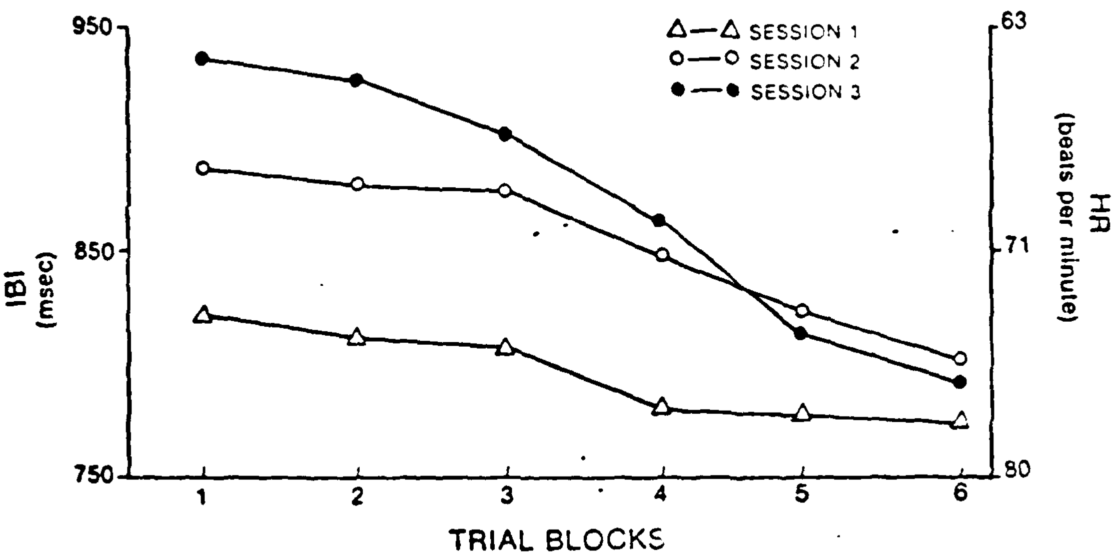

Figure 3b. Mean IBI as a function of trial blocks for the 3 sessions of the carrier task performed alone. Six subjects. Heart rate plotted on right ordinate.

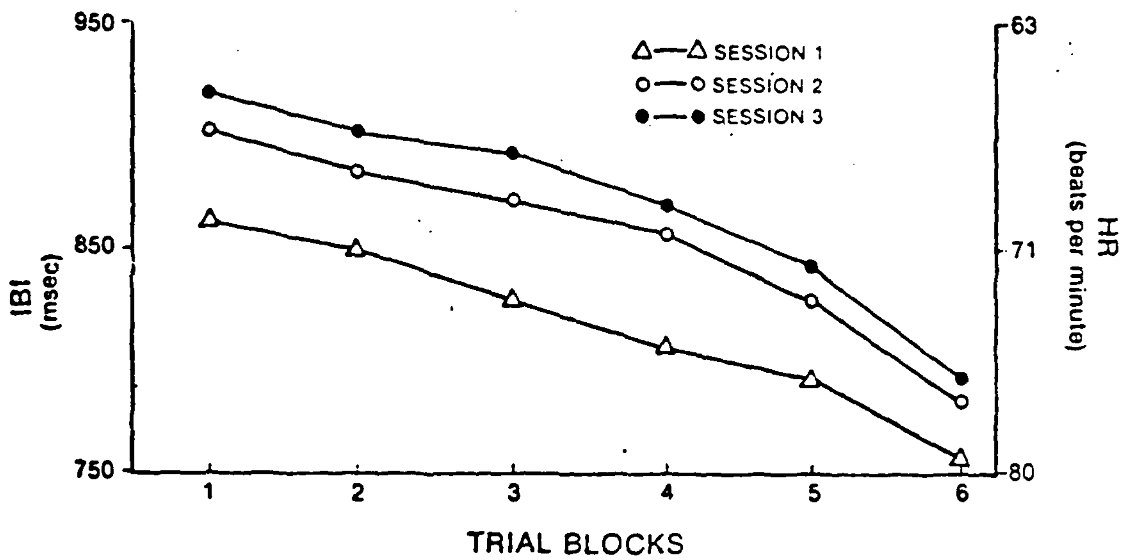

Figure 3c. Mean IBI as a function of trial blocks for the 3 sessions of the carrier task combined with the tone discrimination task. Six subjects. Heart rate plotted on right ordinate.

performance of the tone discrimination task alone, but IBI's decreased over blocks during performance of both the tracking task alone and the tracking plus tone tasks. (For those unfamiliar with IBI's, the corresponding heart rates are plotted on the right ordinate.) The only unexpected result is that IBI's during the tone task alone were shorter than anticipated, given the small workload involved in this task. This might be because this was the first task performed by these subjects in our laboratory and activation levels could have been slightly elevated due to the unfamiliarity of the surroundings.

The ANOVA on the day 2 results revealed significant effects for sessions ($F(2,10)=7.22$, $p < 0.02$) and trial blocks ($F(5,25)=15.73$, $p < 0.001$). Inspection of Figure 3b shows that IBI's decreased dramatically as the subjects flew closer to the carrier landing area. The sessions effect is accounted for by the fact that IBI's increased as a function of practice on the carrier landing task. Finally, the ANOVA on the day 3 results showed a significant blocks effect ($F(5,25)=20.21$, $p < 0.001$), but the sessions effect was not significant ($F(2,10)=3.28$, $p < 0.10$).

From the standpoint of workload assessment, the IBI results agreed with the RT and error results for the secondary task. That is, tone discrimination performance was best on the tone task performed alone, intermediate on the tracking plus tone task, and worst on the carrier landing plus tone task. Similarly, IBI's decreased sharply during performance of the carrier and

carrier plus tone tasks. Note, however, two important differences between the IBI results and the secondary task results: 1) IBI discriminated task workload even when it was not possible to impose a secondary task. It was necessary for the two visuomotor tasks to be performed alone, unencumbered by secondary task demands, in order to evaluate the effects of the combined task situations. On session 2 day 1 (tracking task alone) and all 3 sessions, day 2 (carrier landing task alone), IBI changed systematically as a function of workload. This is particularly clear in the day 2 results in which IBI decreased sharply as the subject flew closer to the carrier landing area.

2) IBI was sensitive to practice effects but secondary task performance was not. On day 2, IBI increased (indicating lower activation) as a function of primary task practice, and this corresponds to the period of rapid learning of the carrier landing task as previously described (see Fig 1b). On day 3, the same pattern of IBI change is seen, although smaller in magnitude, and similarly, primary task performance improved only slightly during this period. (approach scores were stable, but percentage of landings and bolters increased. However, RT's to secondary task stimuli were longer during the second half than during the first half of day 3. This is the opposite of what would be expected since practice and increased mastery would logically mean decreased primary task workload and a corresponding increase (or at least no change) in secondary task performance.

To summarize, IBI was sensitive to long-term decreases in workload due to practice as well as short-term increases in workload represented by final approach to landing. The IBI and secondary task measures of short-term workload agreed well and can be viewed as redundant indicators of the

short-term workload effect, but the IBI results also discriminated long-term changes in workload due to practice while the secondary task measures did not.

Skin conductance

Recall that tone trials were presented each 5 sec and that typical SC responses are long duration and long (e.g., 1 to 3 sec) onset latency phenomena (Edelberg, 1972). Thus, the present paradigm with the relatively short 5 sec inter-stimulus interval was not ideally suited to SC investigation, and as will be seen, there are difficulties with the interpretation of the data.

A software routine searched the SC responses during the time interval between tone trials and found a) the largest SC peak in the time interval, and b) the latency of that peak to the nearest 0.5 sec. The peak could be of either polarity, but the data showed that all peaks were positive-going, indicating increased sympathetic activity. A peak was defined as a change in slope polarity. That is, for a positive-going peak, the slope of the waveform must first be positive, then change to negative. This eliminated trials on which SC changed monotonically during the 5 sec interval and never displayed a definite peak. Such "non-response" trials occurred about 8% of the time, distributed more or less equally across tasks and sessions, and were treated as missing data. This posed no particular problem, since when the SC amplitude and latency data were averaged over blocks, the non-response trials contributed to neither the numerator nor the denominator of the means.

Preliminary analysis showed that there was no effects of response set (respond high or respond low) on SC amplitude or latency, so the data were averaged over this variable. SC amplitude as a function of trial blocks averaged over the 6

subjects is displayed for day 1, 2 and 3 in Figures 4a, 4b, and 4c, respectively. For the day 1 results, (Figure 4a) a repeated measures ANOVA was performed on the three levels of task (tone alone, tracking alone, and tracking plus tone), and 4 levels of trial blocks. Only the trial blocks main effect was significant ($F(3,15) = 8.56$, $p < 0.002$) indicating that, regardless of task, SC amplitude decreased reliably during the 2-min runs.

ANOVA's for the day 2 and day 3 data likewise revealed only trial blocks main effects (for day 2, $F(3,15) = 12.08$, $p < 0.001$, and for day 3, ($F(3,15) = 5.92$, $p < 0.001$) and it is apparent from inspection of Figures 4b and 4c that this represents an increase in SC amplitude during the 2-min runs.

The SC latency functions for the three days are shown in Figures 5a, 5b and 5c. ANOVA's revealed no significant effects for days 1 and 2, but SC latency increased significantly during the 2-min runs on day 3 ($F(3,15) = 10.70$, $p < 0.001$).

A presumption of the design was that SC responses would be elicited by the tone and that one would observe systematic relationships between tone discrimination performance and SC amplitude and latency. However, within-subject correlations (Pearson's r) between RT and SC amplitude, and RT and SC latency revealed r's very near zero. Further, the SC data were sorted separately for trials on which the subjects responded correctly to the tone, and error trials. The mean SC amplitude was slightly larger, and the mean SC latency slightly longer on error trials, but the differences were small and did not approach significance.

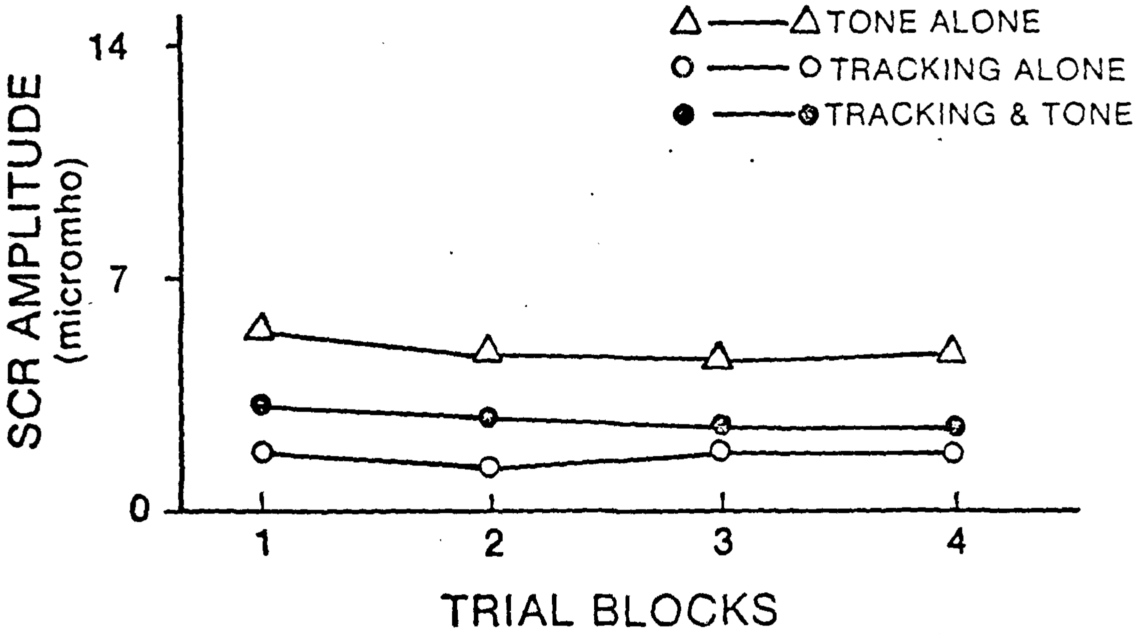

Figure 4a. Skin conductance amplitude as a function of trial blocks for the tone discrimination task performed alone, the tracking task performed alone, and the combined task conditions. Six subjects.

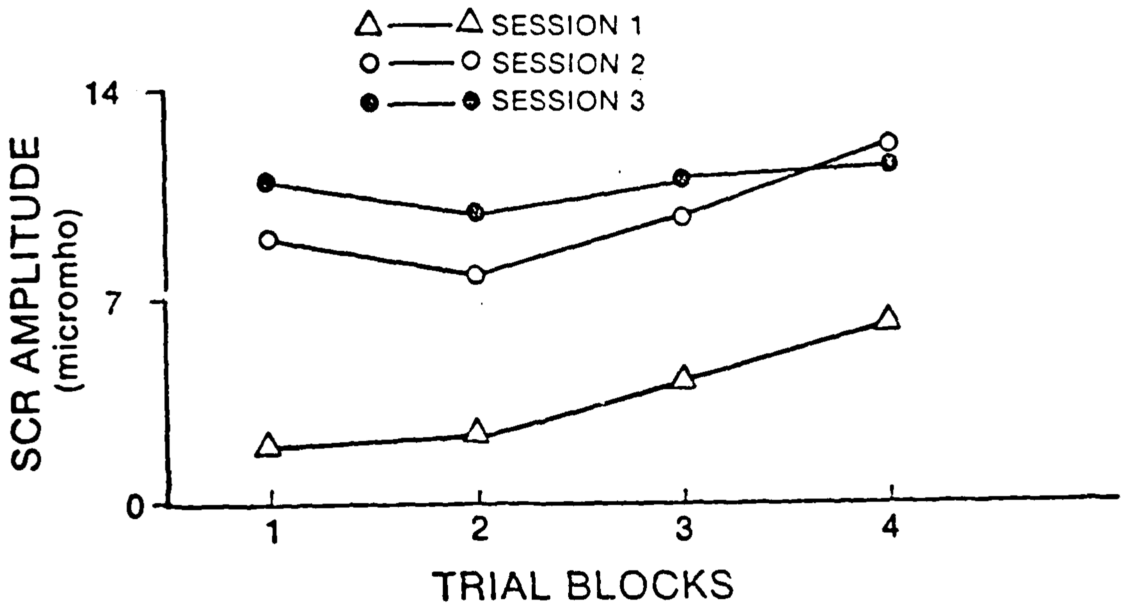

Figure 4b. Skin conductance amplitude as a function of trial blocks for the 3 sessions of the carrier task performed alone. Six subjects.

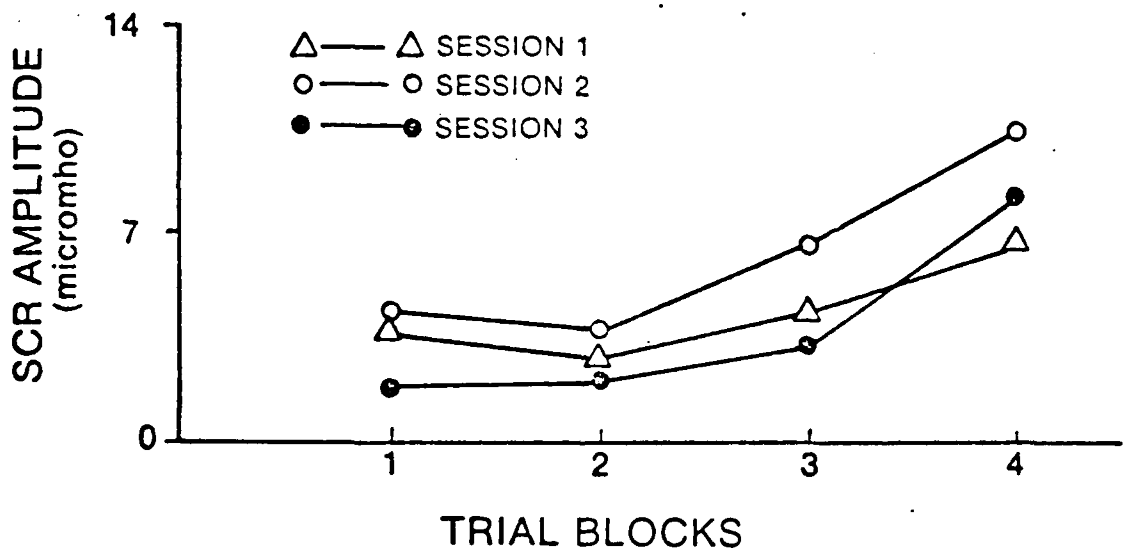

Figure 4c. Skin conductance amplitude as a function of trial blocks for the 3 sessions of the carrier task combined with the tone discrimination task. Six subjects.

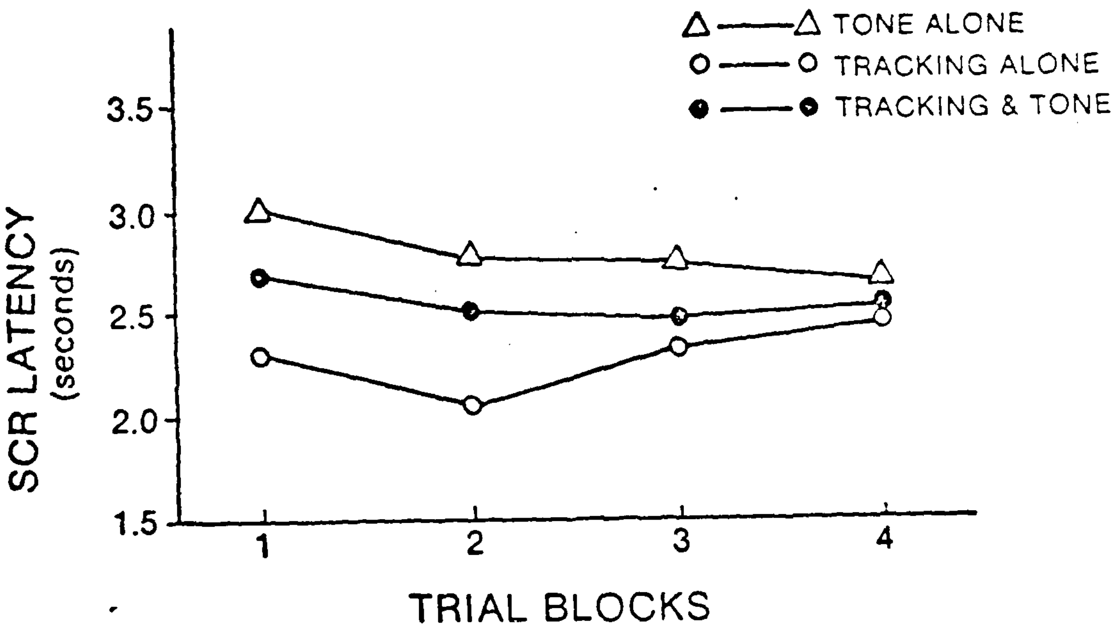

Figure 5a. Latency of skin conductance response as a function of trial blocks for the tone task performed alone, the tracking task performed alone, and the combined task conditions. Mean of 6 subjects.

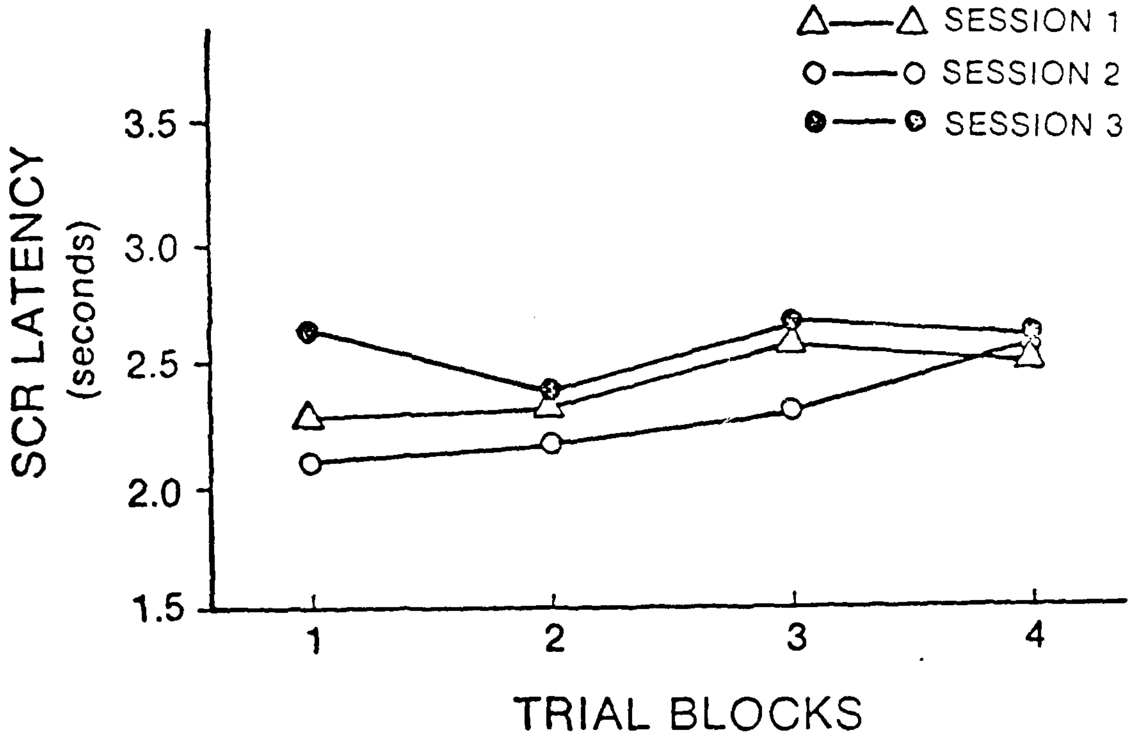

Figure 5b. Latency of skin conductance response as a function of trial blocks for the 3 sessions of the carrier task performed alone. Mean of 6 subjects.

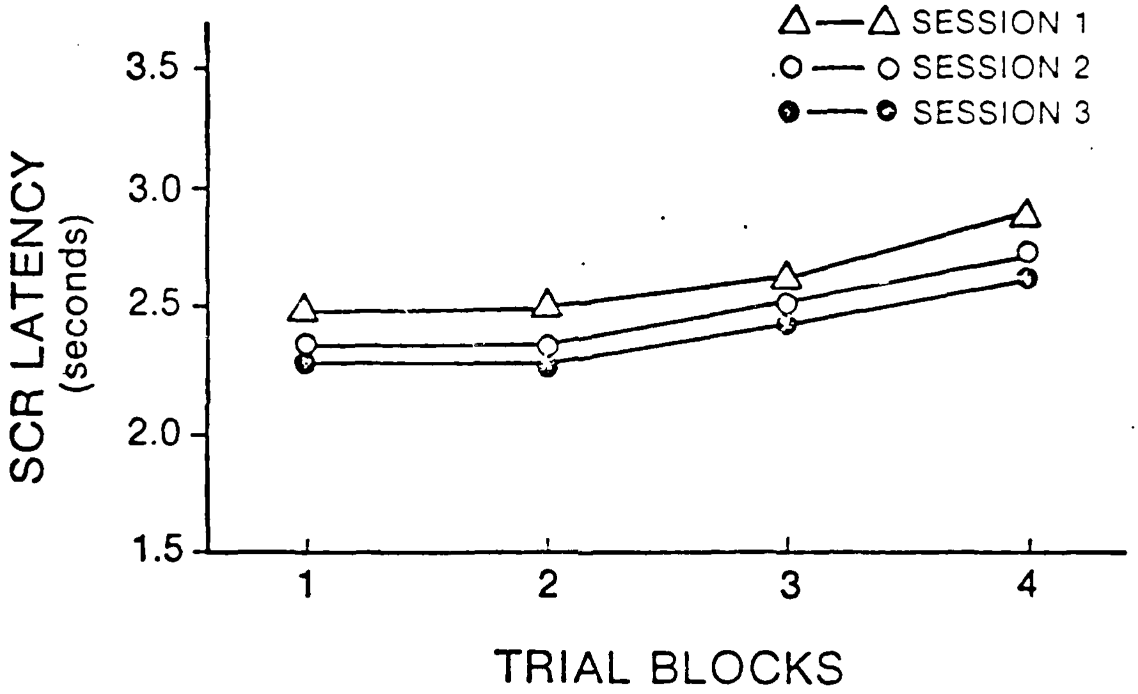

Figure 5c. Latency of skin conductance response as a function of trial blocks for the 3 sessions of the carrier task combined with the tone discrimination task. Mean of 6 subjects.

These results are difficult to interpret. Since correlations between RT and SC amplitude or latency were near zero, it is clear that the functions shown in Figures 4 and 5 are not systematically related to RT. Also, SC amplitudes increased during the 2-min runs on both days 2 and 3, yet on day 2, the tones were irrelevant to the task. Finally, SC responses were not different on error trials compared with correct trials. This lack of relationship between SC responses and tone discrimination performance suggests the hypothesis that some unknown percentage of the SC responses observed in these experiments were not elicited by the tone, but rather, were spontaneous fluctuations related to level of activation. There are several implications of this hypothesis. First, correlations between SC amplitude and IBI should be negative (short IBI's associated with high amplitude SC responses) if both IBI and SC responses are considered measures of activation. This was true in 5 of the 6 subjects although the magnitudes of the correlations were small (-0.1 to -0.3). Secondly, since IBI decreased as subjects flew closer to the carrier, SC amplitude should increase under the same conditions. This is true as shown in Figures 4b and 4c; additionally, SC amplitude was negatively correlated with distance to the carrier (large SC amplitudes associated with short distances to the carrier landing area) in all 6 subjects, but again, the magnitudes of the correlations were small (-0.1 to -0.3). Thirdly, if SC responses are spontaneous and not elicited by the tone, one would expect that SC <u>latency</u> would be related more to SC <u>amplitude</u> than any other variable. This is so because a large amplitude SC response has a long rise and fall time, thus the latency of the peak amplitude would be shifted in the direction of longer latency, given the constraint that SC responses were measured within a 5 sec window in these experiments. Correlational analyses showed that SC latency was positively correlated with SC amplitude in 4 of the 6 subjects (magnitudes of .2 to .3).

Finally, the small decrease in SC amplitudes on day 1, when the more simple tasks were being performed, could be viewed as indicative of the relatively low workload on that day, compared with days 2 and 3 when workload (and SC amplitudes) increased during the phase of final approach to landing.

All factors considered, it appears that SC responses were more indicative of autonomic activation than specific tone stimulus processing and that some unknown number of observations would be contaminated. That is, it is quite possible that a given SC response quantified by our procedure for an arbitrary trial "n" might have actually been an SC response elicited by events occurring late in trial "n-1". This could account for the poor correlations among SC responses and behavioral variables, and would be consistent with the observations that the SC results were similar, but less orderly than the heart rate results.

Event-related potentials (ERP's)

It is standard procedure in modern ERP research to discard records during which the subject blinked the eyes, since eyeblinks can introduce artifacts into the brain wave which can be confused with real ERP's. We employed a stringent criterion which eliminated ERP records whenever any eye movement or blink occurred between the time of stimulus onset and 1000 msec following stimulus onset. As a result, approximately 20% of all ERP records were excluded from the analyses reported below. Since the exclusions were approximately equally distributed over trials, tasks, and conditions, no systematic bias was introduced by discarding data. Although the discard percentage is higher than most published reports, one can have higher than normal confidence that the reported results are essentially free of artifact.

Analysis of the ERP data focused on changes in four prominent components (N1, P2, N2, P3) shown in Fig 6a. Preliminary analysis indicated that none of these components was affected by response set (respond high or respond low), so the reported results are averaged over this variable. Planned comparisons included changes in the brain wave as a function of a) workload (tone, tracking plus tone, and carrier landing plus tone), b) whether the subject was to respond to the tone or inhibit responding to the tone (response vs non-response), c) whether the tone discrimination was hard or easy, d) whether the response to the tone was correct or an error, and e) whether the tone was relevant (tone alone, tracking plus tone, carrier landing plus tone) or irrelevant (tracking alone, carrier landing alone). Latencies and amplitudes of the 4 prominent components were measured by a software routine which accepted, as input parameters, the grand mean of the N1 and P3 latencies. The routine then defined N1 and P3 as the highest amplitude components closest to these means on each trial, and the P2 and N2 as the intervening peak-trough. Amplitude differences of N1P2 and N2P3 were calculated by simple algebraic subtraction. This software routine was developed with considerable effort and careful thought, and we believe it to be a highly reliable method of component identification for single trial data. Validation was accomplished by having two members of the laboratory independently identify the 4 components by visual inspection of hundreds of single trials over a two month period and these judgements were compared against the determinations made by the software routine. Agreement was above 90%.

A repeated measures ANOVA with 3 levels of task (tone alone, tracking plus tone, and carrier landing plus tone) and 2 levels of response (voice response

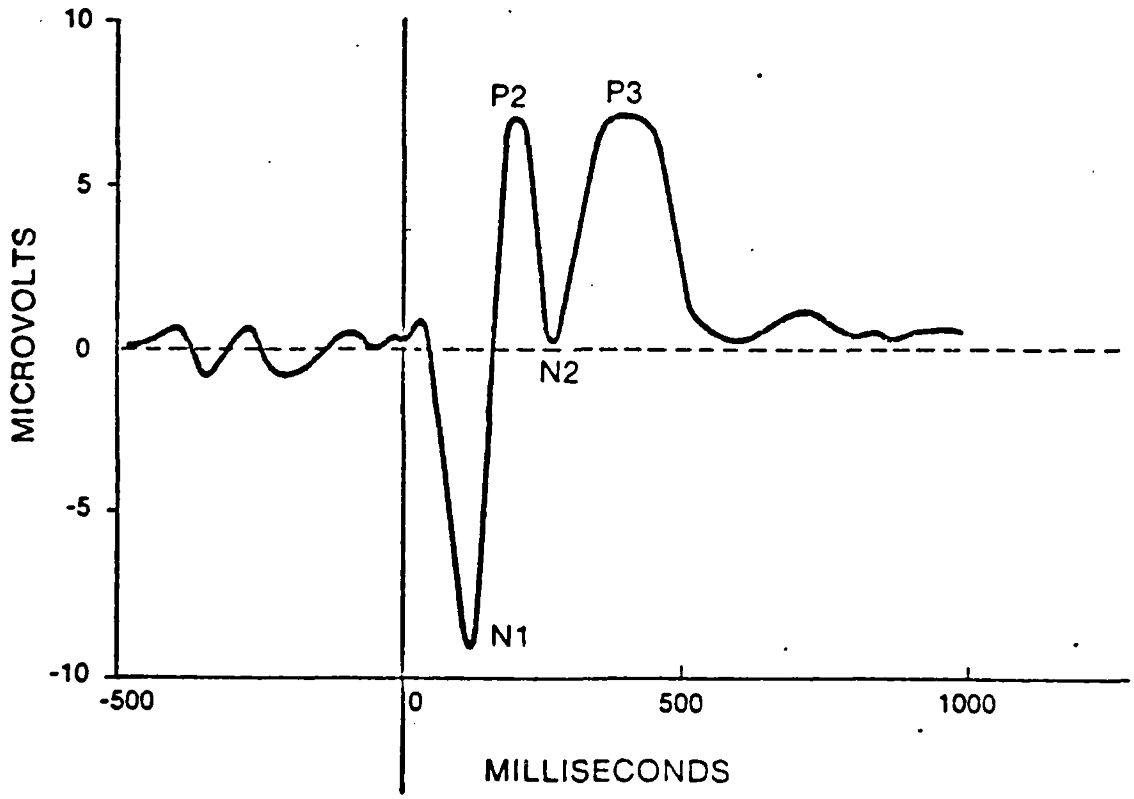

Figure 6a. Typical ERP waveform observed under the conditions of this experiment. Stimulus onset is zero milliseconds on the absissa. Average of 6 subjects times 30 tone presentations.

or no voice response) was performed on the data for the 6 subjects. This analysis was confined to correct responses; errors are treated separately. The main effect of response was not significant, nor was the response by task interaction. This is not to say that the voice response did not cause changes in the brain wave; to the contrary, the voice response was easily identified by a high amplitude, positive slow wave which peaked between 600 and 900 msec following tone onset. The non-significant result simply means that the perturbations in the brain wave caused by the voice response did not contaminate any of the four major components of the ERP which were statistically analyzed.

The task effect was significant for N2 latency (N2L) ($F(2,10)=8.84$, $p < .01$) and the amplitude difference between N2 and P3 ($F(2,10)=17.72$, $p < .001$). These effects, plotted in Figures 6b and 6c, show that N2L increased monotonically as workload increased and N2P3 amplitude decreased monotonically as workload increased. Subsequent ANOVA's were performed pair-wise to determine which task differences accounted for the main effect. Considering first the data from Fig 6b, N2L occurred significantly earlier during the tone alone task than during the tracking plus tone task ($F=14.91$), $p <.02$) or the carrier landing plus tone task ($F=13.05$, $p < .02$). The comparison between tracking plus tone and carrier landing plus tone was not significant. Considering the data presented in Fig 6c, N2P3 amplitude was significantly smaller during the carrier landing plus tone task than during the tracking plus tone task ($F=24.23$, $p < .005$) or the tone task performed alone ($F=63.79$, $p < .001$); the comparison between tone alone and tracking plus tone was not significant. Thus, both N2L and the N2P3 amplitude difference consistently discriminated between the easiest task (tone alone) and the most difficult combination (carrier landing plus tone), but the combination of intermediate difficulty

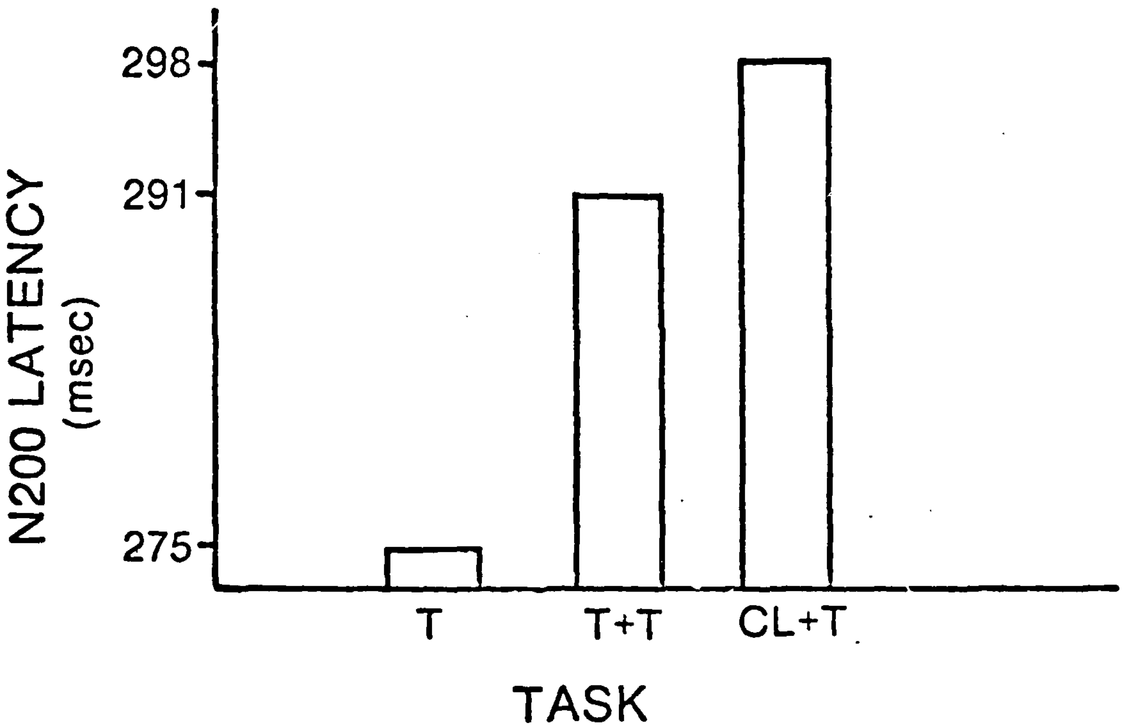

Figure 6b. Latency of the N200 (N2) component of the ERP as a function of task. T= tone discrimination task performed alone, T+T= tone and tracking tasks combined, and CL+T= carrier task and tone task combined. Mean of 6 subjects.

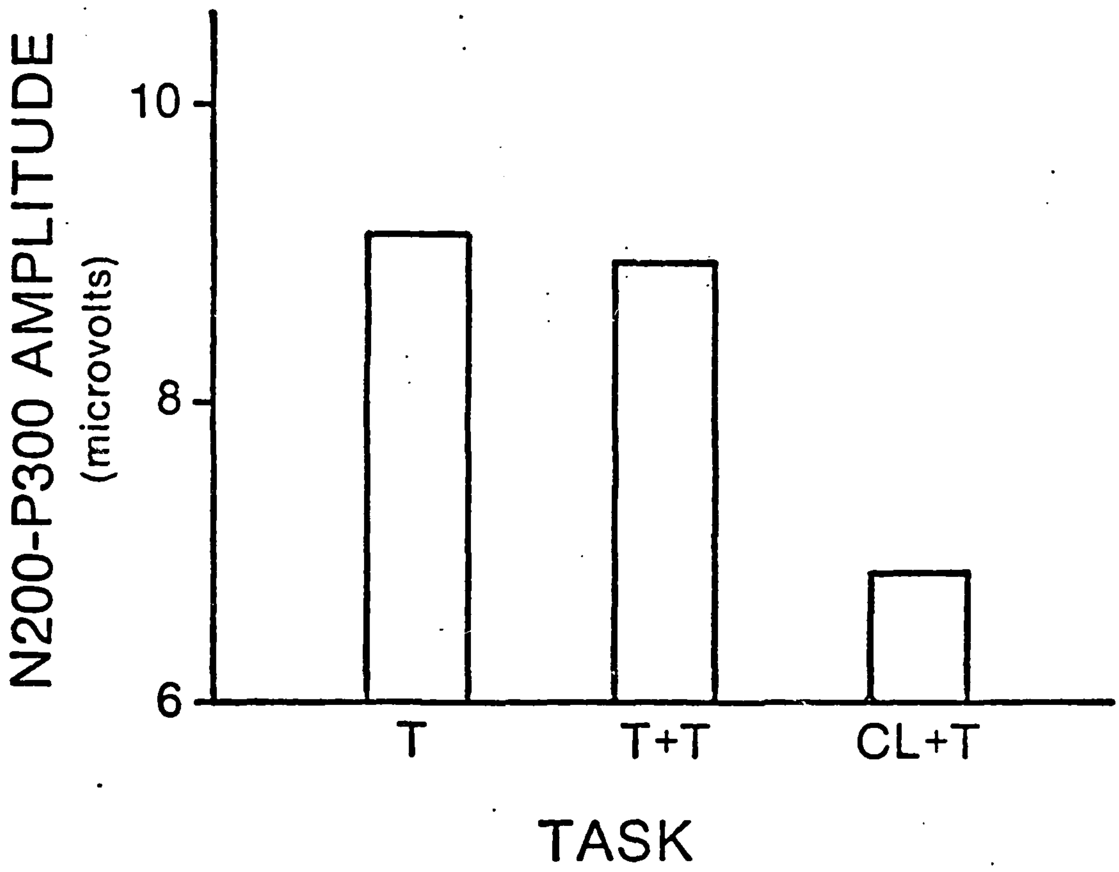

Figure 6c. Trough-peak amplitude of the N2-P3 complex as a function of task. Abbreviations same as Figure 6b.

(tracking plus tone) was not consistently discriminated by differences in the brain wave.

Similar ANOVA's for other planned comparisons revealed the following differences: 1) N1P2 amplitude difference was significantly greater for the easy tone discrimination than for the hard tone discrimination ($F(1,5) = 14.04$, $p < 0.02$); 2) N2L was significantly shorter for correct responses than for errors ($F(1,5) = 9.79$, $p < .05$), and 3) the N2P3 amplitude difference was significantly greater for correct responses than for errors ($F(1,5) = 15.01$, $p < .02$). There were no significant interactions with task, thus these effects were of roughly equal magnitude across the three tasks; 4) N2P3 amplitude difference was significantly larger when the tones were relevant (tone task, tracking plus tone, and carrier landing plus tone) than when the tones were not relevant (tracking alone and carrier landing alone), $F(1,5) = 25.75$, $p < .005$. The means for these effects are summarized in Table 2.

To summarize, N2 latency and N2P3 amplitude discriminated between the lowest and highest workload conditions. Additionally, several of the tone discrimination task variables produced systematic alterations in both early and late ERP components. Specifically, N1P2 amplitude was greater for the easy than for the hard discrimination, N2 latency was shorter and N2P3 amplitude greater for correct responses than for errors, and N2P3 amplitude was greater when the tones were relevant than when the tones were irrelevant.

IV. Discussion: Among-subjects effects

The experiment was designed in a manner which permitted simultaneous evaluation of dual-task methodology and physiological measurement of workload. The physiological measures were chosen to assess a) autonomic activation (heart

TABLE 2

Means of ERP components as functions of discrimination difficulty, whether the response was correct or an error, and whether the tone was task-relevant or task-irrelevant. Amplitudes in microvolts, latencies in milliseconds. Dash entries indicate non-significant differences.

	N1P2 Amplitude	N2 Latency	N2P3 Amplitude
Easy	9.9	---	---
Hard	9.3	---	---
Correct	---	288	8.3
Error	---	294	7.6
Relevant	---	---	8.8
Irrelevant	---	---	6.8

rate, skin conductance) and b) central nervous system activity (event-related potentials). There were two visuomotor tasks (tracking and carrier landing) and one auditory task (tone discrimination). In different conditions, all 3 tasks were performed alone, then the tone task was combined with each of the visuomotor task. In the combined task conditions, subjects were instructed to perform both tasks as well as they could, and the terms "primary" and "secondary" were never used when discussing the tasks with the subjects. However, subjects created their own atmosphere by referring to the carrier landing and tracking tasks as "main" or "major" and clearly viewed the tone discrimination task as less important. Realistically, then, the visuomotor tasks were treated as primary and the tone task as secondary.

The carrier landing task was the most difficult in terms of the complexity of the visual display and the amount of information processing required for high performance. Also, the carrier landing task increased in difficulty as the subjects approached the landing area. This was so because the task simulated the power-on approach (fixed percent power) typical of actual carrier landings, thus accuracy of visual processing, decision making, and joy stick movements became increasingly critical as the distance to the landing area decreased. This situation can be appreciated by imagining the following, analogous task: an automobile and a brick wall are 2 miles apart. The auto has a fixed speed of 60 miles per hour, and no brakes; there is an opening in the wall only a few feet wider than the auto. The task is to drive the auto through the opening. When the task begins, errors in judgement have no great consequence since there is ample time to correct, but as the distance between the auto and wall decreases, even small errors in processing, decision making, and steering wheel

movements become critical. We refer to this as "short-term" workload since it occurred during each 2-min run of the carrier landing task.

Both visuomotor tasks showed large practice effects but, as expected, the tone task did not, since the latter involved only psychophysical judgements requiring nothing more than normal hearing and normal ability to make a response. When the tone task was combined with either of the visuomotor tasks, performance degraded on the tone task but not on the visuomotor tasks. Clearly, subjects treated the tone task as "secondary" and the visuomotor tasks as "primary", ignoring instructions to perform both tasks equally well. It was obvious from the subjects' unsolicited comments that they were highly involved in the carrier landing task and considered the tracking task and the tone tasks dull by comparison.

One purpose of this effort was to compare dual-task methodology and physiological assessment as measures of workload. Dual-task methodology produced the expected results, given that subjects treated the tone task as secondary: Errors and RT's on the tone discrimination task increased when the tone task was combined with the tracking task and increased even more when the tone task was combined with the carrier landing task. Two of the physiological measures, specifically N2 latency and N2P3 amplitude, identified the carrier landing plus tone task as the highest workload condition and the tone task performed alone as the lowest workload condition. Physiological measures were unable to consistently discriminate between the intermediate workload conditions involving the tracking task. Very similar results were reported by Isreal, Wickens, Chesney, and Donchin (1980) who observed reductions in late component (P300) amplitude as a function of increased visual processing load.

The carrier landing task provided interesting within-task comparisons due to the presence of short-term workload increments represented by final approach to landing. These short-term effects were well reflected by both the secondary task results (increased RT's and errors during final approach) and the physiological results (increased heart rate and skin conductance amplitude during final approach). Additionally, the physiological results provided more information than the secondary task results in two instances: a) mean heart rate decreased over sessions reflecting decreased workload due to practice, and b) regardless of whether the secondary task was imposed, heart rate and skin conductance amplitude consistently reflected increases in short-term workload due to final approach to landing. Thus, the physiological measures were superior to secondary task methodology in discriminating long-term decreases in workload due to practice, and the physiological measures were at least as good in discriminating the short-term workload increases due to final approach to landing.

The physiological measures also provided detailed information relevant to secondary task performance. Regardless of whether the tone task was performed alone or in combination with the other tasks, a) the easy tone discrimination evoked larger early components (N1P2 amplitude difference) than the hard tone discrimination, b) correct responses were associated with shorter N2 latencies and larger N2P3 amplitude differences than errors, and c) relevant tones evoked larger N2P3 amplitude differences than when the tones were presented but defined as task-irrelevant. This latter result, that late component amplitude is greater to relevant stimuli than to irrelevant stimuli, is a frequently reported finding (for reviews see Prichard, 1981 or Donchin, Ritter, & McCallum, 1978).

The ERP results are consistent with a simple model which views the early components (N1, P2) as reflecting central nervous system processes involved in stimulus identification (i.e., which tone was presented), while the later components (N2, P3) reflect processes more involved in decision-making (i.e., does the tone require a response?). An important assumption is that amplitudes of components reflect the degree of coherent neural activity where neural activity is viewed as an aggregate of dendritic and synaptic potentials and true action potentials, in short, all of the brain's electrical activity that is recordable within the domain of the surface electrode. Thus, large amplitude potentials indicate a large amount of coherent neural activity and vice-versa. Coherent neural activity is likely to be great when the external stimulus is easy to identify and/or the decision concerning the stimulus is an obvious, easy decision. This is so because one function of the neural process must be a comparison between the incoming stimulus and a memorial representation of the significance or meaning of that stimulus. If the comparison results in an unambiguous decision, large amplitude ERP's will be seen, and if the comparison results in an ambiguous decision, small amplitude ERP's will be seen. This model is consistent with some but not all of the published ERP work and is similar to the target template matching model described by Ford (1978).

It is important to note that the present results show involvement of N2 latency and N2P3 amplitude difference rather than P3 or "P300" amplitude and latency. As Ritter (1978) has pointed out, the N2 has been somewhat neglected in endogenous ERP research, partly because of its smaller amplitude and partly because it is sometimes obscured by the larger P3 which occurs about 100 msec later. We would add to Ritter's observations the following: A large

proportion of late component investigations have utilized the rare stimulus or "odd-ball" paradigm (Donchin, 1980) in which the target stimulus is presented relatively rarely; further, the data have been collected under long time constant recording conditions which passes considerable power in the 1-3 Hz band, and the data have been filtered prior to analysis to remove activity in the alpha band and above. The odd-ball paradigm in conjunction with long time constant recording yields a high amplitude, long duration P3 which does indeed tend to obscure the preceding N2. The filtering further reduces N2 amplitude because it has a period of 100 msec relative to the dominant P3, and the filter acts to remove or reduce activity with a period of 100 msec. In the present experiments, stimuli are equi-probable, a 1 sec time constant was used (to minimize movement artifacts), and the data were analyzed without further filtering. The typical ERP waveform obtained under these conditions (c.f. Fig 6a) shows P2, N2, and P3 of similar amplitude and a P3 duration of about 200 msec. There is no evidence that P3 interferes with N2 under these conditions. Given the present results, Ritter (1978) may be correct in suggesting that the N2 has been neglected and requires more serious investigation. A related problem is why so many ERP studies utilize long time constant recording and filtering of the alpha band frequencies. This produces an ERP waveform heavily biased towards low frequencies, and low frequencies can be present because of movement artifact. The rationale for long time constant recording followed by severe low-pass filtering has never been adequately articulated in the published literature.

The heart rate results are most easily interpreted by an extension of activation theory (Duffy, 1972). By this view, increases in either mental or physical workload increase the metabolic demands on the body and as a result, auto-

nomic activity shifts in the direction of greater sympathetic influence. Signs of increased sympathetic activity include increased heart rate and sweat gland activity as well as several variables not measured in this experiment (e.g., changes in pupil size, increased adrenal corticosteroid output, etc.).

In summary, the among-subject results clearly demonstrated the utility of physiological measures of workload. For the carrier landing task, heart rate was sensitive to both long-term and short-term changes in workload, while dual-task methodology was sensitive to only short-term changes in workload. The experimental design was not well suited to skin conductance measures, and interpretation of this variable was ambiguous in some cases. Late components of the ERP discriminated between the highest and lowest workload conditions, and both early and late ERP components reflected differences in tone discrimination difficulty, accuracy of response, and relevance of the tone stimulus.

Among-subject analyses involve considerable data reduction since variables are averaged into blocks of trials and over subjects. In the next section, attention is turned to within-subject analyses with the intent of providing a finer-grain examination of the relationships among the performance and physiological variables. Recall that a tone trial was presented each 5 sec yielding a 2 sec epoch of ERP data. Mean IBI and IBI change (longest IBI minus the shortest IBI during each 5 sec interval; this is used as a simple measure of heart rate variability) were also computed for each 5 sec interval as well as skin conductance amplitude and latency. Finally, performance measures on the visuomotor tasks were sampled and stored concomitant with each tone presentation along with the RT and the correctness of the response. This permits the use of correlation and regression analysis to examine, for each subject, detailed

relationships among physiology and performance at 5 sec intervals.

V. Results: Within-Subject Effects

Stepwise regression analysis (BMDP P2R, Dixon and Brown, 1979) was performed on the tone discrimination task data with RT as the dependent variable. The predictor variables were IBI, IBI change (IBIC), SCR amplitude (SCRA), SCR latency (SCRL), N1L, P2L, N1P2 amplitude, N2L, P3L, and N2P3 amplitude. For the tracking and carrier landing tasks, multiple performance measures were available. Tracking performance was based on three variables: x and y deviations from the prescribed path, and speed of bug movement (actually, the distance moved during each 5 sec interval). Similarly, flight performance was based on three variables: roll, heading, and vertical speed (rate of descent). In these cases, canonical correlation (BMDP P6M) was used so that the group of variables representing performance could be correlated with the group of physiological variables.

The results for the regression analysis on the tone discrimination task performed alone are summarized in Table 3, along with the rank order of performance for each of the 6 subjects. The performance rank is based only on mean RT. Errors were initially considered, but did not add information. The fastest two subjects had the two lowest error rates, the slowest subject had the highest error rate, and errors for the remaining three subjects were essentially identical. In within-subject analyses, the number of observations varies because different subjects have different numbers of trials with artifacts (blinks, skin conductance responses of zero amplitude, bad IBI record). Also, when the analysis involves RT, error trials are excluded since errors of omission have no RT.

TABLE 3

Rank order of RT performance on the tone task performed alone and summary of regression analysis for each subject. Abbreviations for physiological variables are defined in text. Column heading abbreviations: R= simple (bivariate) correlation of variable indicated with RT; PR= semi-partial correlation of variable indicated with RT after removing effects of variable(s) higher in the list; Cumulative % VAR= the cumulative percentage of RT variance accounted for by the physiological variables selected by the regression analysis. Only significant correlations are reported.

Performance (rank order)	Subject	Variable Name	R	PR	Cumulative % VAR	Number of Observations
		SCRA	.279	.279	8	
1	E	P2L	-.266	.378	14	74
		N2P3	-.226	.452	20	
	B	IBIC	-.263	.263	7	
2		N1L	.261	.371	14	80
3	C	SCRA	.561	.561	31	44
4	H	IBI	.257	.257	6	94
5	F	IBI	-.422	.422	18	66
6	D	N2L	.241	.241	6	77

Examination of Table 3 indicates that autonomic variables generally correlated more strongly with RT than did the ERP variables. The single exception was subject D in which the only significant correlation with RT was N2L, and this relationship accounted for only 6% of the variance. Note that this subject was the poorest performer on the tone discrimination task. Indeed, there is a clear trend for poor performance to be associated with poor correlations between the physiological variables and performance: For the 3 best performers, an average of 21.6% of the variance on RT was accounted for by the physiological measures, while for the 3 worst performers, this figure dropped to 10%. Also, the 2 best performers showed involvement of both autonomic and central nervous system variables with RT performance, while the remaining subjects showed involvement of only one system.

This pattern of results changed when the tone task was combined with the tracking task, as shown in Table 4. There was less involvement of autonomic variables with RT, and a greater involvement of late component ERP amplitude with RT. However, it was still the case that poor RT performance was associated with poor correlations between the physiological measures and RT; indeed, 2 of the 3 worst subjects showed no significant correlations between any of the physiological measures and RT.

Since the subjects were performing the tracking task simultaneously with the tone discrimination task, it is of interest to examine the involvement of the physiological variables with tracking task performance. Canonical correlation analysis was performed for the tracking task performed alone, and for the tracking task combined with the tone discrimination task. The performance variables were the x and y deviations from the prescribed path, measured each 5

TABLE 4

Summary of regression analysis when the tone task was combined with the tracking task. Other particulars same as Table 3. Dashed line indicates no significant correlations with RT.

Performance (rank order)	Subject	Variable Name	R	PR	Cumulative %VAR	Number of Observations
1	B	N2P3	-.446	.446	20	95
2	C	N2P3	-.385	.385	15	125
3	E	IBIC	.393	.393	15	47
4	H	---	---	---	---	62
5	F	---	---	---	---	38
6	D	SKNL	-.314	.314	10	73

sec, and the distance traversed by the bug during that same 5 sec period. Since time was constant (5 sec), this last variable becomes a measure of speed of performance. As was the case with the tone discrimination task, subjects did not adopt a strategy of speed-accuracy trade-off. That is, subjects who earned high tracking scores (low x and y deviation scores) did not accomplish this by creeping around the path at low speed. The rank order of performance on the tracking task was the same if calculated from the x and y deviations alone, or the x and y deviations divided by the speed score.

The results of the canonical correlation analysis for the tracking task performed alone are summarized in Table 5, and for the tracking task performed in conjunction with the tone discrimination task in Table 6.

Examination of the loadings during the learning phase of the tracking task (Table 5) indicates strong individual differences among subjects. For example, for subjects C and E, speed of tracking accounted for most of the performance variance while x deviation was more important to the performance of subjects B and H. On the physiological dimension, autonomic measures uniformly accounted for more variance than the ERP measures, and IBI was the variable most consistently involved. Since tones were irrelevant when the tracking task was performed alone, it is not surprising that the ERP measures correlated less well with performance than the autonomic measures.

Since the canonical variable loadings represent the correlation of each variable with the canonical variable, a workload interpretation suggests itself: The magnitude of the loading reflects the workload represented by the variable. For example, assume that the control aspects of the task were such

TABLE 5

Rank order of tracking performance on the tracking task performed alone, and canonical variable loadings for each of the 6 subjects. For the physiological variables, only the highest three loadings are shown. Significance of the canonical correlation was evaluated by the chi-square test (Dixon and Brown, 1979), and only significant correlations are shown. Xdev= x deviation on tracking task, Ydev= y deviation on tracking task, spd= speed on tracking task.

Performance (rank order)	Subject	Performance Variable	Loading	Physiological Variable	Loading	Number of Observations
		Xdev	.501	IBI	.899	
1	D	Ydev	.448	N2P3	-.349	190
		spd	.331	N1P2	.170	
		spd	.986	IBI	-.831	
2	C	Ydev	.436	SKNA	.436	112
		Xdev	-.111	N1P2	.363	
		spd	.980	SKNA	.943	
3	E	Xdev	.098	IBIC	.259	103
		Ydev	.018	N2P3	.239	
		Xdev	.714	IBI	-.757	
4	B	spd	-.245	SKNA	-.372	170
		Ydev	.206	SKNL	-.287	
		Xdev	.637	IBI	-.803	
5	F	Xdev	.402	N2P3	.392	104
		spd	.395	SKNL	-.251	
		Xdev	.670	SKNL	-.841	
6	H	Ydev	.376	IBI	-.597	139
		spd	.214	IBIC	-.531	

TABLE 6

Canonical correlation results for the tracking task when the latter was performed with the tone discrimination task. Other particulars same as for Table 5.

Performance (rank order)	Subject	Performance Variable	Loading	Physiological Variable	Loading	Number of Observations
1.5	D	---	---	---	---	158
		Xdev	.939	IBI	.605	
1.5	E	Ydev	.562	N1L	-.488	123
		spd	-.284	SKNA	-.382	
		spd	.885	IBI	-.894	
3	B	Ydev	.646	N2P3	.314	225
		Xdev	.369	SKNL	.273	
		Xdev	.676	SKNL	.499	
4	H	spd	-.670	N2P3	-.491	103
		Ydev	-.037	N1L	-.325	
		spd	.990	IBI	-.652	
5	C	Ydev	.134	SKNA	.617	131
		Xdev	.065	N1P2	.384	
6	F	---	---	---	---	69

that it was very easy for the subject to maintain low y deviations. If this were so, y deviation would show a low magnitude loading for all subjects. An even more obvious example would be if another variable, totally unrelated to stick movements and bug control, were entered into the equation. Such a variable would show extremely low loadings with the canonical variable representing the performance dimension. In actuality, the octagon path was used for the tracking task with the goal of equating x and y dimension control difficulty, and the score reported to the subject would be low if he adopted an obvious speed-accuracy trade-off strategy. The most effective strategy, and the one that appeared (by observation) to be adopted eventually by all subjects, was to move the stick in a steady clockwise almost circular path at a rate of about one revolution per 2 sec. Even so, the extent to which x control, y control, or speed contributed to performance was different for different subjects as shown by the loadings.

Table 6 summarizes the canonical correlation results for the tracking task when the latter was performed simultaneously with the tone discrimination task. By now, tracking task performance had stablized at a high level (cf. Fig 1a). Non-significant correlations were obtained for the best and worst performers, and for the remaining 4 subjects, either x deviation or speed showed highest loadings on the performance dimension. For the physiological dimension, autonomic measures again showed higher loadings than the ERP measures.

Considering Tables 3 through 6 together, the following pattern emerges: When the tone discrimination task or the tracking task was performed alone, relationships among autonomic measures and performance were stronger than relationships among central nervous system measures and performance. This was

true regardless of whether the performance variable was RT or tracking. When the two tasks were combined, there was a shift such that autonomic variables remained associated with tracking performance while late component amplitude of the ERP became more associated with RT. Another major point is that individual differences become apparent with this type of analysis. Subjects F and H were consistently below median performance in all four performance evaluations (RT on tone alone task, RT on combined task, tracking accuracy on tracking alone task, and tracking accuracy on combined task). Subject E was consistently above median performance in all four performance evaluations, and subjects C and B were above median performance in three of the four evaluations. Subject D was uniquely interesting since he was consistently worst on the tone discrimination task and consistently best on the tracking task. The consistently poor performers (subjects F and H) showed a paucity of physiological involvement with the RT task, and one of these subjects (subject F) showed no significant relationships among the tracking performance variables and physiological measures in the dual-task condition. This could lead to a tentative hypothesis that poor relationships among physiological variables and performance variables are indicative of poor performance, with the exception of subject D who displayed the unique pattern of poorest performance on the tone discrimination task, best performance on the tracking task, but no significant correlations among the tracking performance variables and the physiological variables.

Tables 7, 8, and 9 summarize the results for the carrier landing task performed alone and in combination with the tone discrimination task. Performance ranks for the carrier landing task are based on both the approach score and number of landings. Also, only the first session of the carrier alone and carrier plus tone tasks are considered, since these are most comparable to the single session of tracking alone and tracking plus tone tasks. Performance measures

TABLE 7

Rank order of performance and canonical loadings for each subject on the carrier landing task performed alone. Abbreviations defined in text.

Performance (rank order)	Subject	Performance Variable	Loading	Physiological Variable	Loading	Number of Observations
		Roll	.954	IBI	.940	
1	C	HED	.755	SKNA	-.278	327
		VSPD	.496	IBIC	.154	
2	E	---	---	---	---	306
3	D	---	---	---	---	279
		HED	.858	SKNA	.850	
4	H	ROLL	.813	IBI	.467	318
		VSPD	.710	IBIC	.191	
		ROLL	.911	IBI	.985	
5	B	HED	.830	SKNA	-.392	353
		VSPD	.611	IBIC	.013	
		HED	.858	SKNA	.850	
6	F	ROLL	.813	IBI	.467	318
		VSPD	.710	IBIC	.191	

TABLE 8

Rank order of performance and canonical loadings for each subject on the carrier landing task when the latter was performed simultaneously with the tone discrimination task. Abbreviations defined in text.

Performance (rank order)	Subject	Performance Variable	Loading	Physiological Variable	Loading	Number of Observations
		ROLL	.691	SKNA	.812	
1	E	VSPD	.321	IBI	-.645	85
		HED	.173	P2L	.271	
2	C	---	---	---	---	105
		ROLL	.732	IBI	-.704	
3	D	VSPD	-.674	SKNA	.673	175
		HED	-.448	IBIC	-.585	
4	H	---	---	---	---	212
		HED	.546	IBI	.794	
5	B	ROLL	-.524	P2L	.291	200
		VSPD	-.465	N2P3	.284	
6	F	---	---	---	---	98

TABLE 9

Rank order of RT performance and summary of regression analysis for the tone discrimination task when the latter was performed simultaneously with the carrier landing task. Other particulars same as Table 3.

Performance (rank order)	Subject	Variable Name	R	PR	Cumulative % VAR	Number of Observations
		N1P2	-.307	.307	9	
1	B	N2P3	-.296	.426	19	81
		P2L	-.256	.478	23	
		N2P3	-.296	.296	9	
2	H	SKNA	.241	.372	14	77
3	C	---	---	---	---	33
4	F	---	---	---	---	36
5	E	---	---	---	---	33
6	D	SKNA	-.352	.352	12	77

are deviations of roll (ROLL), heading (HED), and vertical speed (VSPD) from the ideal flight path.

Considering first the data in Table 7, it appears that vertical speed (descent rate) was the least difficult flight parameter to control during this first session which represented early learning of the carrier landing task, and only autonomic nervous system variables were involved in flight performance. For 2 of the 3 best performers, no significant canonical correlation was obtained. Table 8 shows the analogous results for the first session of the combined task condition; at this stage, flight performance was quite stable for the group (c.f. Figure 1b), although there were still large individual differences. There is no longer a clear pattern identifying which flight control parameter seems most or least important to the performance dimension, as might be expected during this later stage of learning when the subjects are "fine-tuning" their control of the aircraft. Autonomic variables are still more important in the relationship between physiology and performance than central nervous system variables. Finally, only 3 of the 6 subjects showed significant canonical correlations, and 2 of these 3 were below median performers. The results of the regression analysis on RT for the combined task condition are summarized in Table 9. Only 3 of the 6 subjects show significant relationships, and the better performers show stronger relationships than the poorer performers. As was the case when the tone task was combined with the tracking task, there is a greater involvement of central nervous system variables with RT performance, relative to when the tone task was performed alone.

Considering just the tone discrimination task, whether performed alone or in

conjunction with either of the visual tasks, there was a clear tendency for poor RT performance to be associated with poor correlations among the physiological measures and RT. This can be seen by examination of Tables 3, 4, and 9; the 3 subjects whose RT's were faster than the grand median RT (the "better" performers) generally showed a higher cumulative percent variance (that is, a stronger relationship between RT and the physiological measures) than the 3 subjects with RT's below the grand median (the "poorer" performers). Further, this relationship changed as the individual subject's performance changed, so is not, apparently, an artifact of some biological tendency on the part of an individual to show consistently low or consistently high physiological responsivity. For example, the performance of subject E was above the median in Tables 3 and 4, and significant correlations were seen between RT and some of the physiological measures. However, when his performance dropped (Table 9), so did the correlations between RT and the physiological measures. Additional examples are clear from examination of the Tables. To determine whether a similar phenomenon was present in the visual task results, additional stepwise regression analyses were performed. Since this analysis permits only one dependent variable, the results of the canonical analysis were used to select the most appropriate dependent variable, specifically, the variable which was the most important single variable in determining the performance dimension. Examination of Table 5 shows that this would bx z-deviation for subject D, speed for subject C, and so on. The results showed the same tendency seen in the tone discrimination results: The 3 better performers on the tracking task performed alone (subjects D, C, and E, cf. Table 5) showed cumulative percent variance figures of 47%, 4%, and 32%, respectively, while the 3 poorer performers (B, F, H) showed figures of 12%, no significant correlations, and 7%, respectively. Finally, the same approach was applied to the other 3 visual

task conditions (tracking plus tone, carrier landing alone, and carrier landing plus tone), and the trend was present in all conditions. Exceptions were subject D in the tracking plus tone and carrier landing alone conditions, subject E in the carrier landing alone condition, and subject C in the carrier landing plus tone condition. These cases showed no significant correlations between performance and the physiological measures yet their performance in these conditions was above median performance.

VI. Discussion: Within-Subject Effects

Analyses were performed on the data for each subject separately in order to search for consistencies in strategies used by the subjects, and to closely examine the relationships among the physiological variables and the performance measures. Examination of the performance rank orders for each subject in Tables 3 through 9 shows individual differences in the subjects' abilities to perform the tasks alone and in combination. For example, subject E was the best performer on the tone discrimination task when performed alone, but his tone discrimination performance degraded sharply in the dual-task conditions as he maintained high performance on the visual tasks. This is indicative of the strategy which drops the priority of the secondary task in order to do well on the primary task. Subject B shows nearly the reverse strategy; he consistently performed well on the tone discrimination task even in the dual task conditions, his tracking performance was moderately good, but his carrier landing task performance was poor. This is indicative of a strategy which aims to maintain high performance on the simpler task while tolerating poorer performance on the more difficult tasks. The performance of subject C was most indicative of one who followed instructions carefully; he performed consistently and moderately well on the tone discrimination task and moderately well

on the visual tasks, suggesting an honest attempt at sharing effort equally between the two tasks when performed simultaneously. Subject D consistently treated the tone discrimination task as a low priority, and performed well to moderately well on the two visual tasks. Finally, subject F was the most consistently poor performer in all task conditions. Whether this represents a motivational problem or true lack of ability cannot be determined from the present data.

One might have expected one or two subjects to be consistently superior to the other subjects in all tasks and all conditions. This would be likely to occur in a truly randomly selected sample from a large, heterogenous population. Recall, however, that the subjects in these experiments were very homogenous (affiliates of the Air Force ROTC program) and not randomly selected, but recruited because of their interest in flying and flight-related research. This method of subject selection was used to circumvent the problems of having randomly selected subjects in the experiment who might be poor performers for a variety of unspecifiable reasons (i.e., hostile attitude toward military research, total lack of interest in flying, unusually poor vision or hearing). With this in mind, two points become clear: First, the worst performer in this experiment might well rank close to the best performer in a group of truly randomly selected subjects, and secondly, in a homogenous population of subjects, one should not expect to observe extreme differences among subjects.

The analyses of the relationships between physiological measures and performance yielded several interesting findings. When any of the three tasks were performed alone, correlations between autonomic nervous system variables and performance were higher than correlations between central nervous system

variables and performance, but when the tone task was combined with either of the visuomotor tasks, there was a clear tendency for the autonomic variables to become more associated with visuomotor task performance. It can be hypothesized that this "autonomic shift" is indicative of the amount of effort subjects invested in the various tasks: Any task performed alone requires no sharing of effort, and performance is most closely related to the state of the autonomic nervous system. But when two tasks are performed simultaneously, sharing of effort is required, and performance on the more demanding task (in the present case, the visuomotor tasks) will be more closely associated with the state of the autonomic nervous system than performance on the less demanding task (tone discrimination). This hypothesis is consistent with activation theory, but obviously requires testing against independent sets of data. Indeed, the hypothesis in its present form cannot totally account for all of the present data, for example, subjects E and D performed well on the carrier landing task alone but showed non-significant canonical correlations. However, the hypothesis could be consistent with cases in which extremely high or extremely low performance is associated with non-significant correlations among physiological and performance variables simply because lack of physiological involvement with performance would be indicative of low effort. Low effort could equally describe the superior performer (because the task is easy for him) and the grossly inferior performer (because he is not trying, or is so overwhelmed by the task demands that he is expending effort in ways that were not quantified in this experiment.) Examples of extremely high and extremely low performers showing non-significant relationships among physiological and performance variables are apparent in Table 6 and less extreme cases are apparent in Table 9.

One of the statistical techniques used for the within-subject analysis was canonical correlation, which is a rarely used statistic in psychological research. In effect, canonical correlation permits correlating one group of variables which are presumed to form a dimension, with another group of variables which are presumed to form a different dimension. In the present experiments, visuomotor task performance was characterized by 3 variables and physiological state was characterized by several measures of autonomic and central nervous system activity. The technique appeared valuable since it was possible to characterize which of the performance measures contributed most to the performance dimension for each individual subject. By examination of the loadings, one can make statements about how each of the variables are related to the dimension. For example, control of vertical speed in the carrier landing task consistently showed the smallest loading of the three control parameters (cf. Table 7), permitting the inference that vertical speed was the easiest to control. Similarly, IBI change consistently showed small loadings, suggesting that heart rate variability is a relatively non-sensitive measure in designs of this type. Other investigators might wish to make greater use of canonical correlation since, at least in the present experiments, the information provided was valuable.

VII. Summary and Conclusions

1. Dual task methodology and physiological assessment of workload were compared in a design involving a tone discrimination task at two levels of difficulty and two visuomotor tasks (tracking and carrier landing). Although subjects were instructed to perform both tasks equally well in the combined task conditions, most treated the tone discrimination task as "secondary" (low priority). Given this is true, the results were typical

of secondary task paradigms: Tone discrimination performance degraded when the tone task was combined with the simple visuomotor task (tracking) and degraded even more when the tone task was combined with the complex visuomotor task (carrier landing). One anomalous result was that tone discrimination performance became worse as a function of visuomotor task practice; apparently, most subjects became more involved in the visuomotor tasks and correspondingly less involved in the secondary task as a function of practice on the visuomotor tasks.

2. For the carrier landing task, both short term and long term workload changes were pronounced. Short term workload refers to the phenomenon that, during each 2 minute run, workload increased steadily as the subject flew closer to the carrier landing area. This was paralleled by signs of increased autonomic activation (increased heart rate and skin conductance amplitude) and, in the dual task condition, decreased secondary task performance. Long term workload changes refer to the phenomenon that, with practice, subjects displayed increased mastery of the carrier landing task which logically implies decreased workload. These practice, or training effects were paralleled by substantial decreases in mean heart rate during the early (first three) sessions when the carrier landing task was necessarily performed alone, unencumbered by secondary task demands. The secondary task was imposed during the final three sessions; small additional decreases in long term workload were observed, paralleled by small additional decreases in mean heart rate. However, as described above, secondary task performance became worse as primary task performance became better. In short, heart rate accurately described changes in both short term and long term workload, while secondary task performance accurately

described changes only in short term workload.

3. Late components of the vertex ERP (N2 latency and N2P3 amplitude) discriminated among-task workload. Specifically, N2 latency increased and N2P3 amplitude decreased monotonically as workload increased across conditions. Secondary task performance also discriminated these workload differences, and the effect was statistically more robust than for the late component ERP effect. However, the ERP data provided information concerning tone task stimulus processing which could not be provided by the reaction time and error scores. Specifically, N1P2 amplitude was greater for the easy tone discrimination than for the difficult tone discrimination; also, N2 latency was shorter, and N2P3 amplitude greater for correct responses than for errors. This suggests that central nervous system processes concerned with stimulus identification can occur at least as early as P2 (about 200 msec following stimulus onset), and decision-making processes can begin as early as N2 (about 275 msec following stimulus onset). Finally, in agreement with other published reports, late component amplitude was greater when the tone was task relevant than when the tone was task irrelevant.

4. Of particular relevance to flight training research was the observation that mean heart rate decreased as mastery of the carrier landing task increased. This suggests that heart rate might be an excellent index of training effectiveness. Similarly, the observation that the heart rate increases during final approach were resistant to training effects suggests that heart rate can be used to pinpoint other maneuvers in other situations which are consistent and stable high workload periods. Finally, it is suggested that the ERP measures could be used to describe the efficacy of

auditory information processing in training environments where the normal task of the pilot is to process auditory information while simultaneously controlling the aircraft.

5. It is concluded that physiological assessment has several advantages over dual task methodology in the study of operator workload, training effects, and performance. One major advantage is the non-intrusive nature of the physiological approach. Additionally, the use of secondary task performance to infer changes in primary task workload can be of questionable validity because the experimenter cannot adequately control individual subject strategies regarding the relative priorities of the primary and secondary tasks. Finally, the combined use of autonomic and ERP measures can provide more detailed descriptions of short and long term workload changes and central nervous system information processing than is possible with dual task methodology.

VIII. References

Ahern, S., and Beatty, J. Pupillary responses during information processing vary with scholastic aptitude test scores. Science, 1979, 205, 1289-1292.

Dixon, W.J. and Brown, M.B. Biomedical Computer Programs P-Series, University of California Press, 1979.

Donchin, E. Surprise!...Surprise? Psychophysiology, 1981, 18:5, 493-513.

Donchin, E., Ritter, & McCallum, C. Cognitive Psychophysiology: The endogenous components of the ERP. Chapter in Event Related Brain Potentials in Man. Callaway, Tueting, & Koslow (eds), Academic Press, New York, 1978.

Duffy, E. Activation. In N.S. Greenfield and R.A. Sternback (eds) Handbook of Psychophysiology, New York: Holt, Rinehart, and Winston, 1972.

Edleberg, R. Electrical activity of the skin: Its measurement and uses in psychophysiology. In N.S. Greenfield and R.A. Sternback (eds) Handbook of Psychophysiology, New York: Holt, Rinehart, and Winston, 1972.

Ford, J.M. Does P300 reflect template match/mismatch? In D. Otto (ed), Multidisciplinary perspectives in event-related brain potential research. EPA-600/9-77-043, U.S. Government Printing Office, Washington, D.C., 1978.

Isreal, J.B., Wickens, C.D., Chesney, G.L., and Donchin, E. The event-related potential as an index of display-monitoring workload. Human Factors, 22:2, 211-224, 1980.

Kahneman, D., Tursky, B., Shapiro, D., and Crider, A. Pupillary, heart rate, and skin conductance changes during a mental task. Journal of Experimehtal Psychology, 1969, 79:1, 164-167.

Lang, P.J., Gatchel, R.J. and Simons, R.F. Electro-cortical and cardiac rate correlates of psychophysical judgement. Psychophysiology, 1975, 12:6, 649-655.

Lindholm, E., Ruppel, M., and Buckland, G.H. Attention and task complexity as indicated by physiological indices. AFHRL TR-79-47, 1979.

McLeod, P. Does probe RT measure central processing demands? Quarterly Journal of Experimental Psychology, 1978, 30, 83-89.

Pritchard, W.S. Psychophysiology of P300. Psychological Bulletin, 1981, 89, 506-540.

Ritter, W. Latency of event-related potentials and reaction time. In D. Otto (ed), Multidisciplinary perspectives in event-related brain potential research. EPA-600/9-77-043, U.S. Government Printing Office, Washington, D.C., 1978.

Roman, J.A., Older, H., and Jones, W.L. Flight research program: VII. Medical monitoring of Navy carrier pilots in combat. Aerospace Medicine, 1967, 38, 133-139.

Wierwille, W.W. and Williges, R.C. Survey and analysis of operator workload assessment techniques. Report #S-78-101, Systemetries, Inc., P.O. Drawer S, Blacksburg, VA.

ATE

LME